

Pada modul praktikum ini, kita akan mempelajari bagaimana cara menginstalasi peralatan untuk big data dimulai dengan hadoop. Singkatnya hadoop merupakan alat untuk memproses data berukuran besar atau big data menjadi beberapa bagian atau disebut cluster untuk referensinya kalian bisa baca di sini.

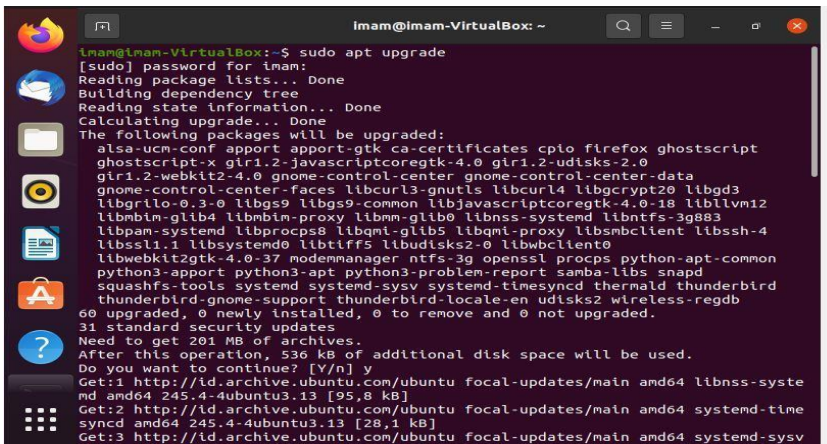
Sebelum memulai praktikum perlu diperhatikan bahwa modul ini akan menggunakan sistem operasi berbasis Linux khususnya **Ubuntu 20.04 LTS (Menggunakan Virtual Box)** agar dapat stabil menggunakan tools big data nantinya. Apabila menggunakan sistem operasi lain diperbolehkan dan dapat disesuaikan dengan modul ini karena kemungkinan pengaturannya atau menjalankan toolsnya akan berbeda.

Langsung saja kita coba bagaimana cara menginstalasi hadoop di **Ubuntu 20.04 LTS**, berikut langkah-langkahnya :

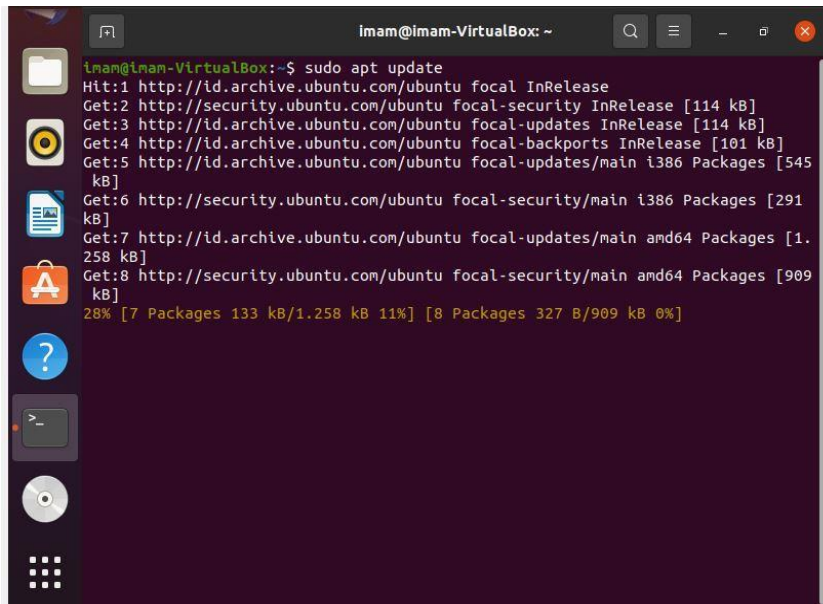
A. Pendahuluan

1. Pastikan sistem kalian sudah tersambung ke internet untuk melakukan update software yang terbaru supaya menghindari error saat instalasi.

```
sudo apt upgrade
```



```
sudo apt update
```



```
imam@imam-VirtualBox: ~  
imam@imam-VirtualBox:~$ sudo apt update  
Hit:1 http://id.archive.ubuntu.com/ubuntu focal InRelease  
Get:2 http://security.ubuntu.com/ubuntu focal-security InRelease [114 kB]  
Get:3 http://id.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]  
Get:4 http://id.archive.ubuntu.com/ubuntu focal-backports InRelease [101 kB]  
Get:5 http://id.archive.ubuntu.com/ubuntu focal-updates/main i386 Packages [545  
kB]  
Get:6 http://security.ubuntu.com/ubuntu focal-security/main i386 Packages [291  
kB]  
Get:7 http://id.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages [1.  
258 kB]  
Get:8 http://security.ubuntu.com/ubuntu focal-security/main amd64 Packages [909  
kB]  
28% [7 Packages 133 kB/1.258 kB 11%] [8 Packages 327 B/909 kB 0%]
```

2. Instalasi JDK 8 dikarenakan hadoop membutuhkannya dalam pengaturannya.

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

```
mazharrasyad@aspire:~$ sudo apt install openjdk-8-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
openjdk-8-jdk is already the newest version (8u282-b08-0ubuntu1~20.04).  
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
```

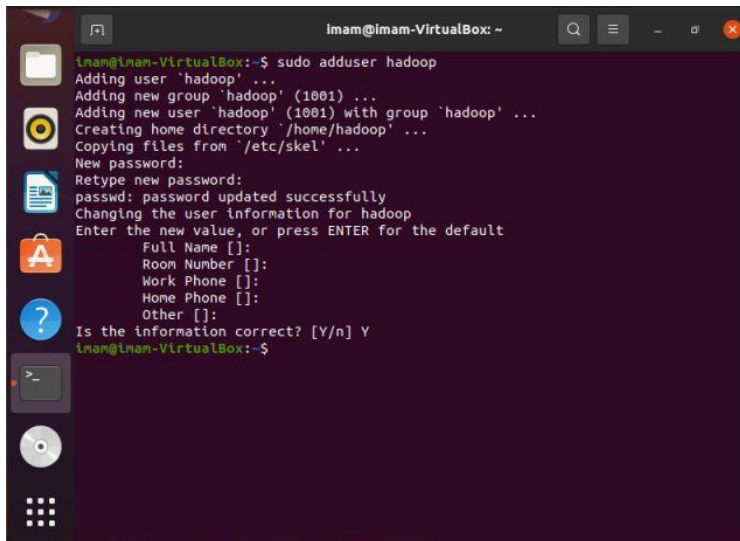
3. Jika sudah selesai instalasi JDK 8, cek kembali dengan perintah berikut apakah sudah berhasil atau belum.

```
java -version
```

```
mazharrasyad@aspire:~$ java -version  
openjdk version "1.8.0_282"  
OpenJDK Runtime Environment (build 1.8.0_282-8u282-b08-0ubuntu1~20.04-b08)  
OpenJDK 64-Bit Server VM (build 25.282-b08, mixed mode)
```

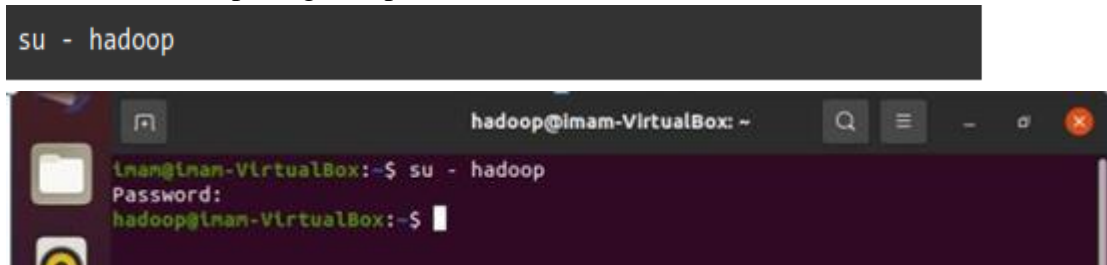
4. Membuat user baru khusus untuk menggunakan hadoop di komputer kita, username yang diisikan yaitu hadoop untuk password disarankan diisikan kata hadoop, dan data diri tidak perlu diisi namun cukup tekan Enter saja hingga diakhiri dengan mengetikkan huruf Y.

```
sudo adduser hadoop
```



```
imam@imam-VirtualBox: ~  
$ sudo adduser hadoop  
Adding user 'hadoop' ...  
Adding new group 'hadoop' (1001) ...  
Adding new user 'hadoop' (1001) with group 'hadoop' ...  
Creating home directory '/home/hadoop' ...  
Copying files from '/etc/skel' ...  
New password:  
Retype new password:  
passwd: password updated successfully  
Changing the user information for hadoop  
Enter the new value, or press ENTER for the default  
Full Name []:  
Room Number []:  
Work Phone []:  
Home Phone []:  
Other []:  
Is the information correct? [Y/n] Y  
imam@imam-VirtualBox:~$
```

5. Jika sudah berhasil membuat user hadoop, coba lakukan perintah berikut untuk masuk ke dalam user hadoop dengan superuser.



```
su - hadoop  
  
imam@imam-VirtualBox:~$ su - hadoop  
Password:  
hadoop@imam-VirtualBox:~$
```

6. Selanjutnya kita membutuhkan konfigurasi SSH karena sistem hadoop membutuhkannya dalam mengelola setiap nodenya. Sehingga setiap user yang mengaksesnya harus terlebih dahulu melakukan konfigurasi SSH untuk user hadoop. Langsung saja kita lakukan generate public key SSH di user hadoop untuk autentikasi dengan mengetikkan perintah berikut dan kemudian tekan Enter saja sampai selesai.



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

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ 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:  
SHA256:1prgf02HASLZ2Y6NXB8Bt2TwkKl4kB/neOUr9f9EWu4 hadoop@imam-VirtualBox  
The key's randomart image is:  
+----[RSA 3072]----+  
  +.+++=.  
  =+**+.  
  *8+0.  
  ..0*+  
  .S...+00|  
  .00+0=|  
  .000.+|  
  . .0.+|  
  ..E|  
+----[SHA256]-----+  
hadoop@imam-VirtualBox:~$
```

7. Setelah hasil generate public key SSH didapatkan dalam file id_rsa.pub maka kita perlu menyimpannya dalam file khusus supaya bersifat read only yaitu file authorized_keys dan mengatur permissionnya.

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
chmod 640 ~/.ssh/authorized_keys
```

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ ls .ssh/  
id_rsa id_rsa.pub  
hadoop@imam-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
hadoop@imam-VirtualBox:~$ ls .ssh/  
authorized_keys id_rsa id_rsa.pub  
hadoop@imam-VirtualBox:~$ chmod 640 ~/.ssh/authorized_keys  
hadoop@imam-VirtualBox:~$
```

8. Terakhir dalam konfigurasi SSHnya yaitu mencobanya dengan perintah berikut

```
ssh localhost
```

```
hadoop@lmam-VirtualBox: ~  
hadoop@lmam-VirtualBox:~$ ssh localhost  
The authenticity of host 'localhost (127.0.0.1)' can't be established.  
ECDSA key fingerprint is SHA256:7J8HNMT91E0y+u48nCA4Ek2uAdSTX25AIEf7at0jpw0.  
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes  
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.  
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-36-generic x86_64)  
  
 * Documentation:  https://help.ubuntu.com  
 * Management:    https://landscape.canonical.com  
 * Support:       https://ubuntu.com/advantage  
  
103 updates can be applied immediately.  
50 of these updates are standard security updates.  
To see these additional updates run: apt list --upgradable  
  
Your Hardware Enablement Stack (HWE) is supported until April 2025.  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.  
hadoop@lmam-VirtualBox:~$
```

Jika terjadi hal berikut, maka pastikan kalian telah menginstalasi SSH terlebih dahulu.

```
hadoop@aspire:~$ ssh localhost  
ssh: connect to host localhost port 22: Connection refused  
hadoop@aspire:~$
```

Caranya cukup mengetikkan perintah berikut namun harus dilakukan di superuser dan bukan user hadoop, oleh karena itu kita perlu keluar dari user tersebut.

```
exit  
sudo apt install ssh
```

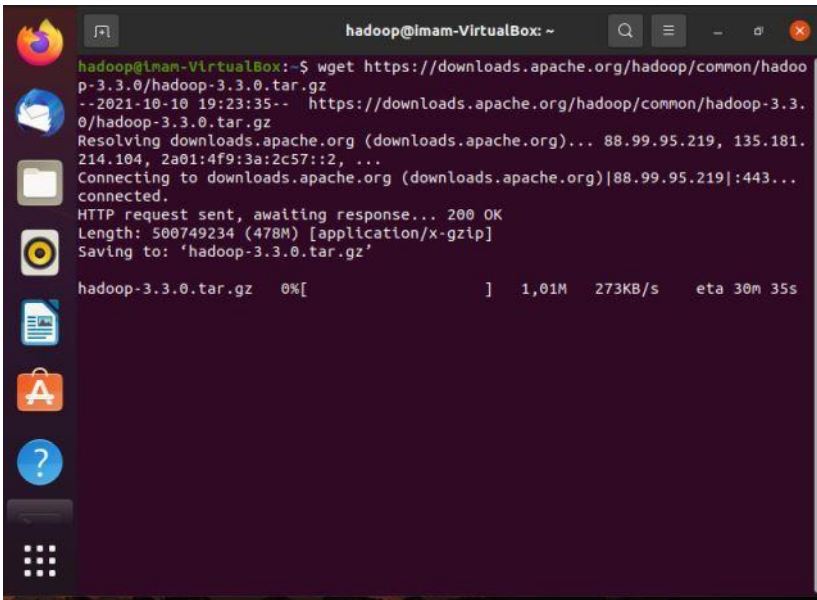
Jika sudah berhasil instalasi SSH, maka kalian dapat mencobanya lagi dengan terlebih dahulu masuk ke user hadoop dan ketikkan perintah yang sama tanpa harus generate public key karena sudah kita buat sebelumnya.

```
hadoop@lmam-VirtualBox: ~  
hadoop@lmam-VirtualBox:~$ ssh localhost  
The authenticity of host 'localhost (127.0.0.1)' can't be established.  
ECDSA key fingerprint is SHA256:7J8HNMT91E0y+u48nCA4Ek2uAdSTX25AIEf7at0jpw0.  
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes  
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.  
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-36-generic x86_64)  
  
 * Documentation:  https://help.ubuntu.com  
 * Management:    https://landscape.canonical.com  
 * Support:       https://ubuntu.com/advantage  
  
103 updates can be applied immediately.  
50 of these updates are standard security updates.  
To see these additional updates run: apt list --upgradable  
  
Your Hardware Enablement Stack (HWE) is supported until April 2025.  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.  
hadoop@lmam-VirtualBox:~$
```


B. Instalasi Hadoop

1. Tahap selanjutnya download package hadoop berikut.

```
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
```



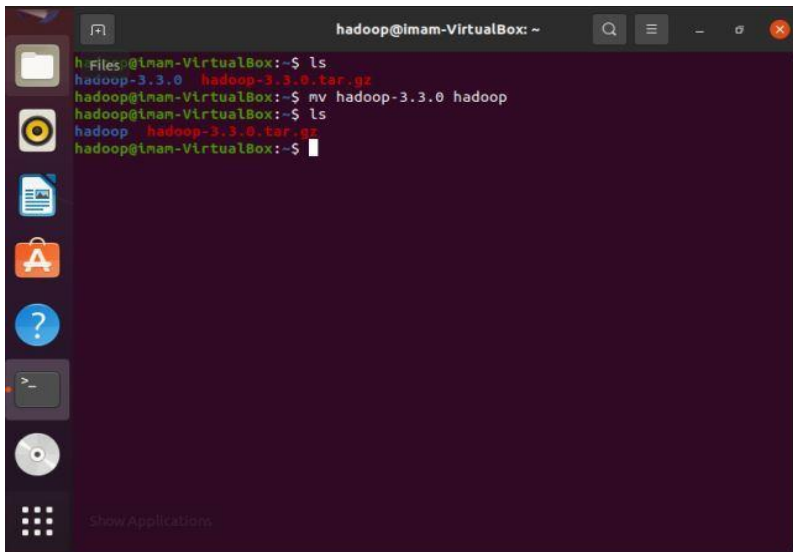
```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz  
--2021-10-10 19:23:35-- https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz  
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.214.104, 2a01:4f9:3a:2c57::2, ...  
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 500749234 (478M) [application/x-gzip]  
Saving to: 'hadoop-3.3.0.tar.gz'  
  
hadoop-3.3.0.tar.gz  0%[          ] 1,01M  273KB/s  eta 30m 35s
```

2. Lakukan extract ketika download sudah selesai maka akan muncul folder hadoop

```
tar -xvzf hadoop-3.3.0.tar.gz
```

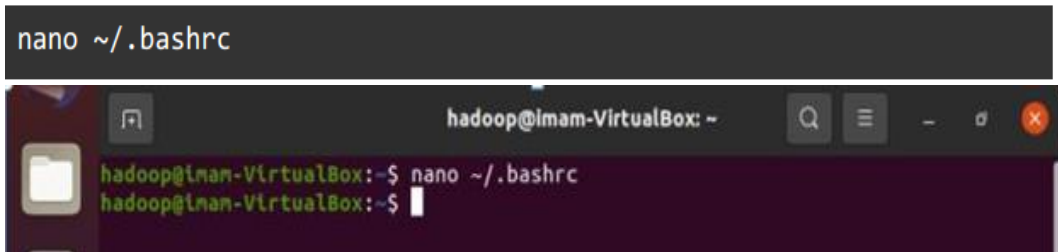
3. Ubah nama folder yang sebelumnya kita extract menjadi hanya hadoop agar mudah.

```
mv hadoop-3.3.0 hadoop
```



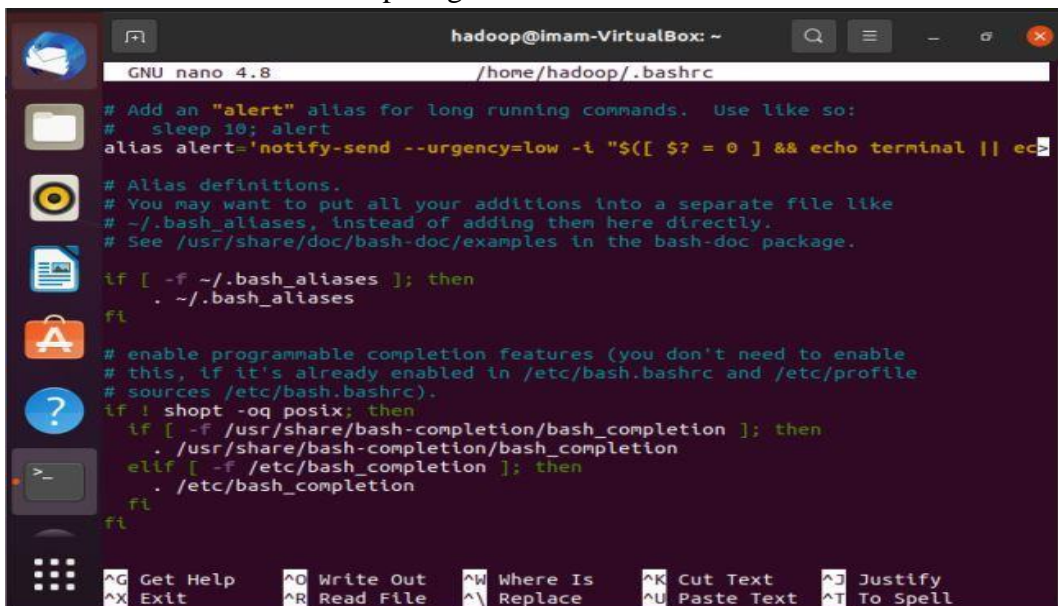
```
hadoop@imam-VirtualBox: ~  
h Files@imam-VirtualBox:~$ ls  
hadoop-3.3.0  hadoop-3.3.0.tar.gz  
hadoop@imam-VirtualBox:~$ mv hadoop-3.3.0 hadoop  
hadoop@imam-VirtualBox:~$ ls  
hadoop  hadoop-3.3.0.tar.gz  
hadoop@imam-VirtualBox:~$
```

4. Lakukan konfigurasi pada file bashrc untuk menginisiasi environment hadoop ke dalam sistem.



```
nano ~/.bashrc  
  
hadoop@imam-VirtualBox:~$ nano ~/.bashrc  
hadoop@imam-VirtualBox:~$
```

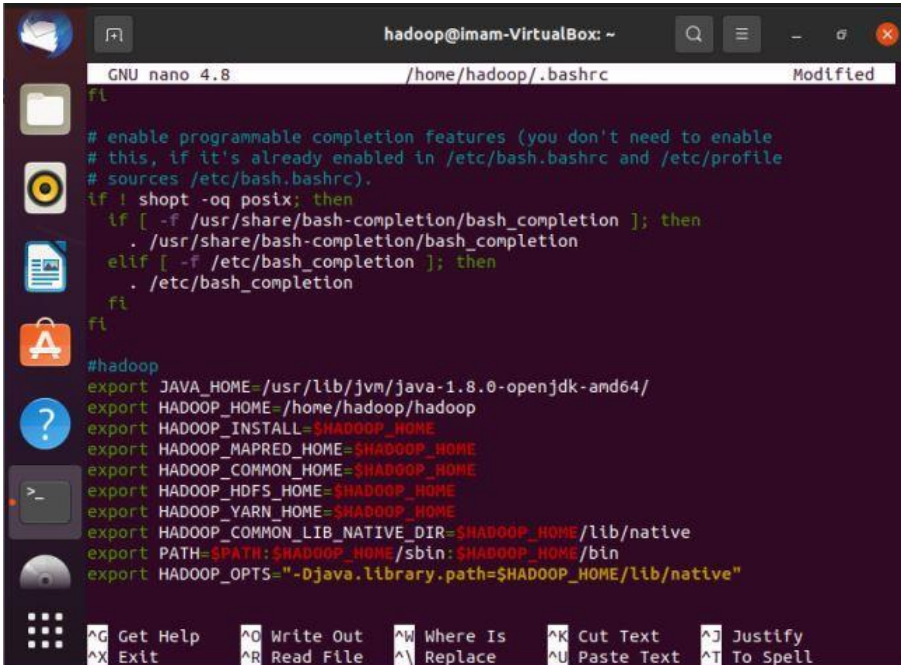
Pada file tersebut arahkan ke paling bawah untuk kita tambahkan kode di akhir baris.



```
GNU nano 4.8 /home/hadoop/.bashrc  
  
# Add an "alert" alias for long running commands.  Use like so:  
# sleep 10; alert  
alias alert='notify-send --urgency=low -i "$([ $? = 0 ] && echo terminal || echo $0)" "Script finished: $@"'  
  
# Alias definitions.  
# You may want to put all your additions into a separate file like  
# ~/.bash_aliases, instead of adding them here directly.  
# See /usr/share/doc/bash-doc/examples in the bash-doc package.  
  
if [ -f ~/.bash_aliases ]; then  
  . ~/.bash_aliases  
fi  
  
# enable programmable completion features (you don't need to enable  
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile  
# sources /etc/bash.bashrc).  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
  
^G Get Help      ^O Write Out    ^W Where Is     ^K Cut Text     ^J Justify  
^X Exit          ^R Read File    ^_ Replace      ^U Paste Text   ^T To Spell
```

Masukkan perintah berikut ke dalam file bashrc di paling bawah baris dan jika sudah tekan CTRL + X untuk keluar dan tekan huruf Y lalu Enter untuk menyimpannya.

```
# Hadoop
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64/
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 HADOOP_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"
```



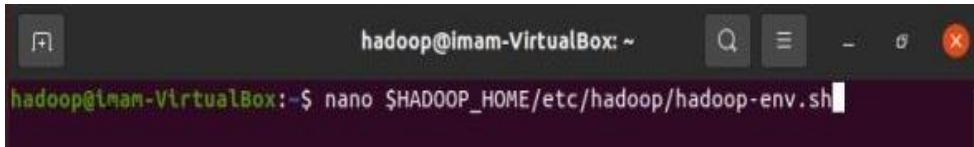
5. Setelah selesai tinggal kita aktifkan dengan perintah berikut.

```
source ~/.bashrc
```

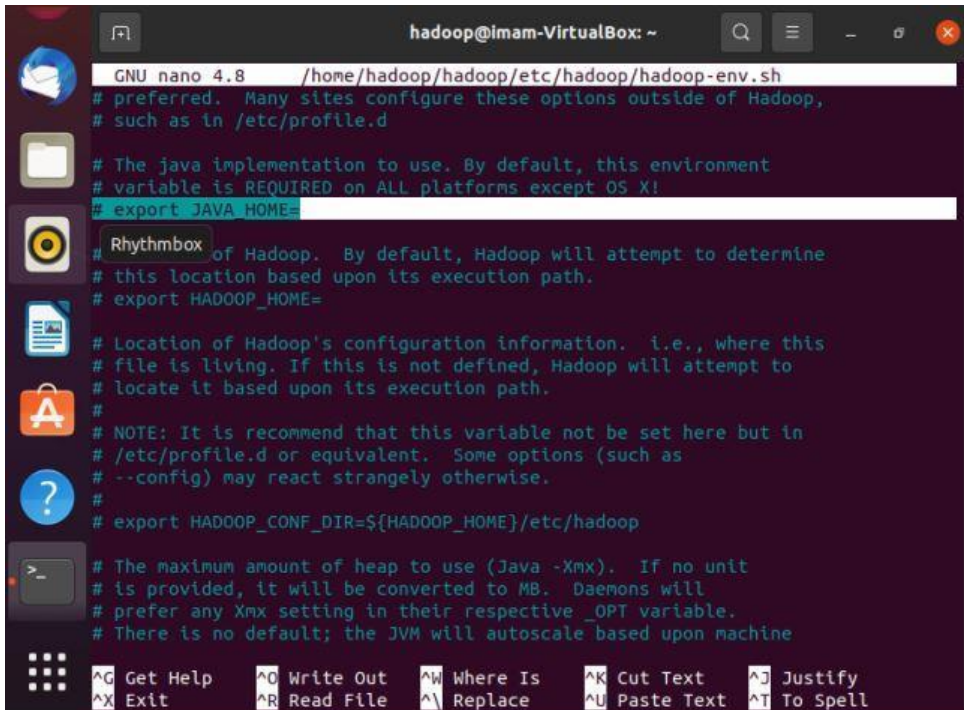


6. Kemudian kita konfigurasi JDK yang sebelumnya kita instal di file hadoop-env.sh, perhatikan perintah \$HADOOP_HOME telah kita inisiasi di file bashrc sebelumnya. Sehingga apabila tidak muncul maka pastikan kembali telah menambahkan environmentnya.


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

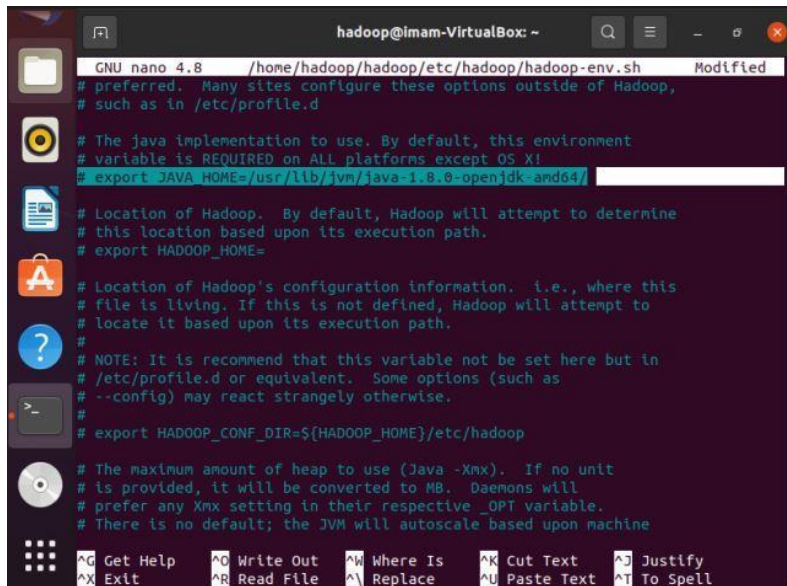


Setelah membuka file tersebut maka cari sintak `#export JAVA_HOME =`



Ubah dengan menggunakan kode berikut dan kemudian simpan.

```
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64/
```



```
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh Modified
# preferred. Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
# export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64/

# Location of Hadoop. By default, Hadoop will attempt to determine
# this location based upon its execution path.
# export HADOOP_HOME=

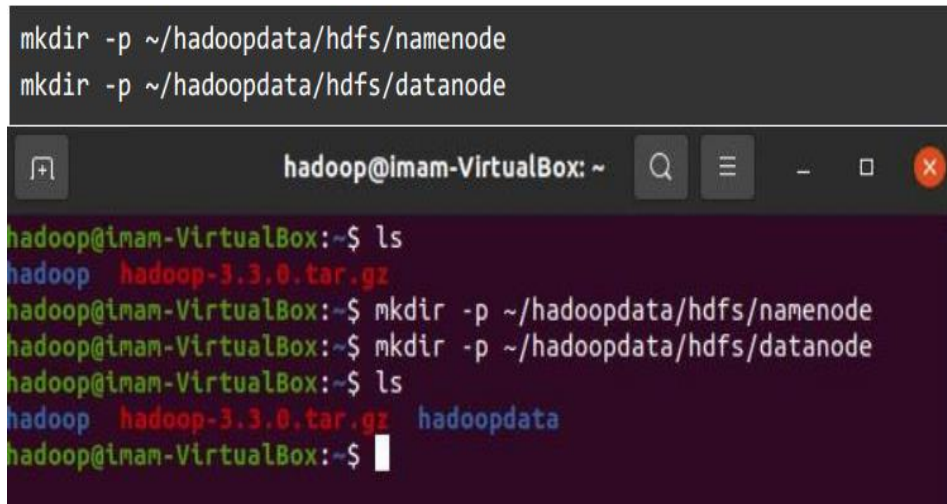
# Location of Hadoop's configuration information. i.e., where this
# file is living. If this is not defined, Hadoop will attempt to
# locate it based upon its execution path.
#
# NOTE: It is recommend that this variable not be set here but in
# /etc/profile.d or equivalent. Some options (such as
# --config) may react strangely otherwise.
#
# export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop

# The maximum amount of heap to use (Java -Xmx). If no unit
# is provided, it will be converted to MB. Daemons will
# prefer any Xmx setting in their respective _OPT variable.
# There is no default; the JVM will autoscale based upon machine

Get Help Write Out Where Is Cut Text Justify
Exit Read File Replace Paste Text To Spell
```

C. Konfigurasi Hadoop

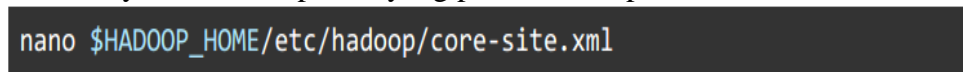
1. Tahap sebelumnya kita telah berhasil download hadoop dan sedikit menginisiasinya di dalam sistem kita. Namun, kita masih perlu melakukan beberapa konfigurasi untuk dapat menggunakannya dimulai dari membuat direktori hadoopdata..



```
mkdir -p ~/hadoopdata/hdfs/namenode
mkdir -p ~/hadoopdata/hdfs/datanode

hadoop@imam-VirtualBox:~$ ls
hadoop hadoop-3.3.0.tar.gz
hadoop@imam-VirtualBox:~$ mkdir -p ~/hadoopdata/hdfs/namenode
hadoop@imam-VirtualBox:~$ mkdir -p ~/hadoopdata/hdfs/datanode
hadoop@imam-VirtualBox:~$ ls
hadoop hadoop-3.3.0.tar.gz hadoopdata
hadoop@imam-VirtualBox:~$
```

2. Berikutnya ada beberapa file yang perlu diubah pertama core-site.xml.



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

Fokus pada bagian `<configuration>` `</configuration>` karena akan kita isi perintah berikut dan kemudian simpan.

```
hadoop@imam-VirtualBox: ~
/home/hadoop/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
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
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
..Ubuntu Software

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

<configuration>
</configuration>
```

[Read 20 lines]

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://127.0.0.1:9000</value>
  </property>
</configuration>
```

```
hadoop@imam-VirtualBox: ~
/home/hadoop/hadoop/etc/hadoop/core-site.xml Modified
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
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
LibreOffice Writer

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

<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://127.0.0.1:9000</value>
  </property>
</configuration>
```

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell

3. Selanjutnya file hdfs-site.xml

```
hadoop@aspire: ~
hadoop@aspire:~$ nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

Sama seperti file sebelumnya, fokus ke bagian `<configuration>` `</configuration>` dan masukkan perintah berikut.

```
<configuration>

    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>

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

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

    </property>
</configuration>
```

```
Thunderbird Mail
hadoop@Imam-VirtualBox: ~
/home/hadoop/hadoop/etc/hadoop/hdfs-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

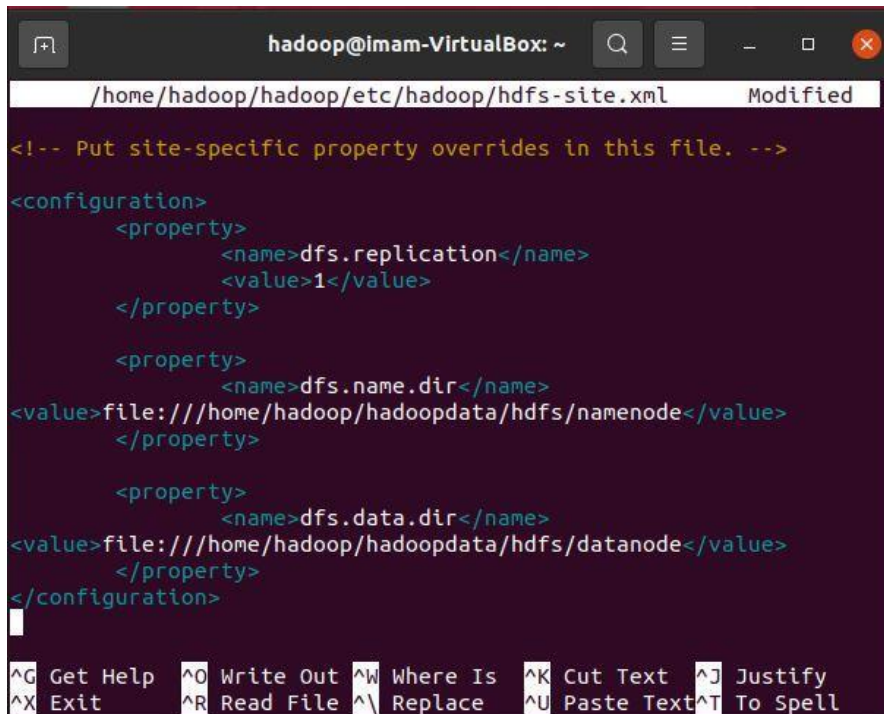
    http://www.apache.org/licenses/LICENSE-2.0

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

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

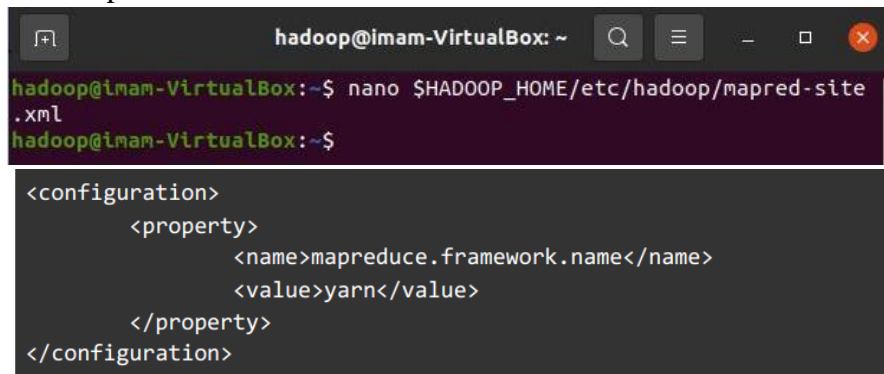
<configuration>

</configuration>
```

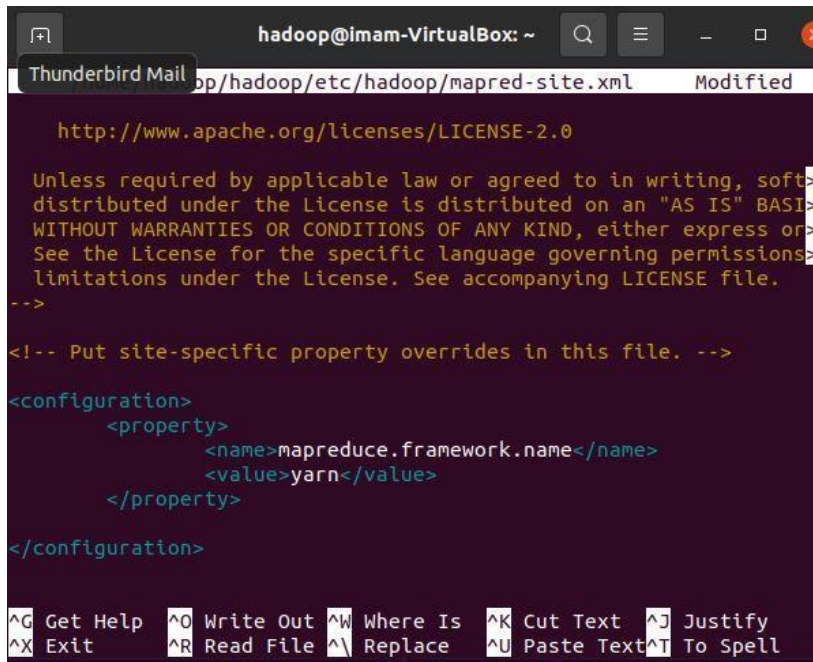


```
hadoop@imam-VirtualBox: ~  
/home/hadoop/hadoop/etc/hadoop/hdfs-site.xml Modified  
  
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
  <property>  
    <name>dfs.replication</name>  
    <value>1</value>  
  </property>  
  
  <property>  
    <name>dfs.name.dir</name>  
<value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>  
  </property>  
  
  <property>  
    <name>dfs.data.dir</name>  
<value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>  
  </property>  
</configuration>  
  
^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify  
^X Exit      ^R Read File  ^\ Replace   ^U Paste Text ^T To Spell
```

4. File mapred-site.xml

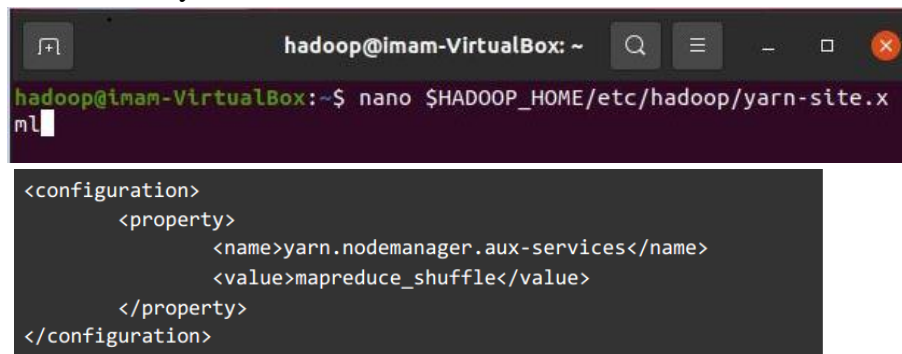


```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ nano $HADOOP_HOME/etc/hadoop/mapred-site.xml  
hadoop@imam-VirtualBox:~$  
  
<configuration>  
  <property>  
    <name>mapreduce.framework.name</name>  
    <value>yarn</value>  
  </property>  
</configuration>
```

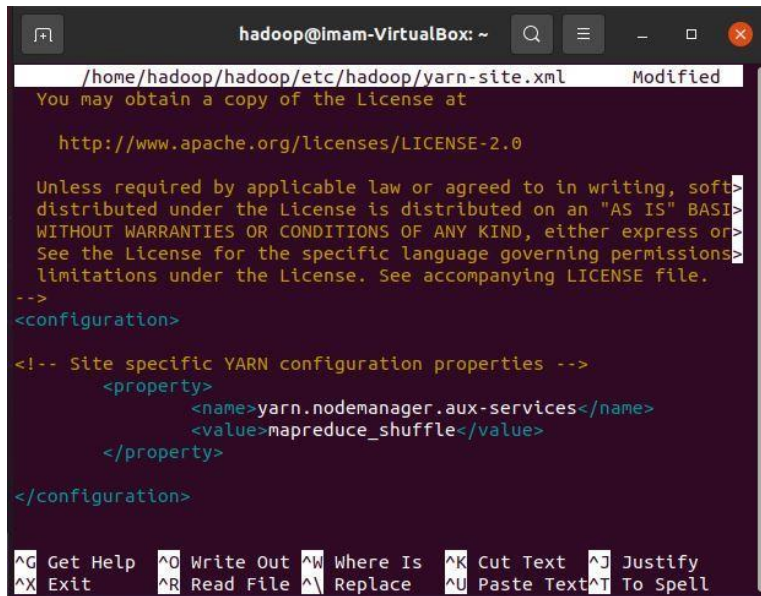



```
hadoop@imam-VirtualBox: ~  
Thunderbird Mail | /etc/hadoop/mapred-site.xml | Modified  
  
http://www.apache.org/licenses/LICENSE-2.0  
  
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR 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.  
-->  
  
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
  <property>  
    <name>mapreduce.framework.name</name>  
    <value>yarn</value>  
  </property>  
</configuration>  
  
^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify  
^X Exit      ^R Read File  ^\ Replace   ^U Paste Text ^T To Spell
```

5. Terakhir file yarn-site.xml



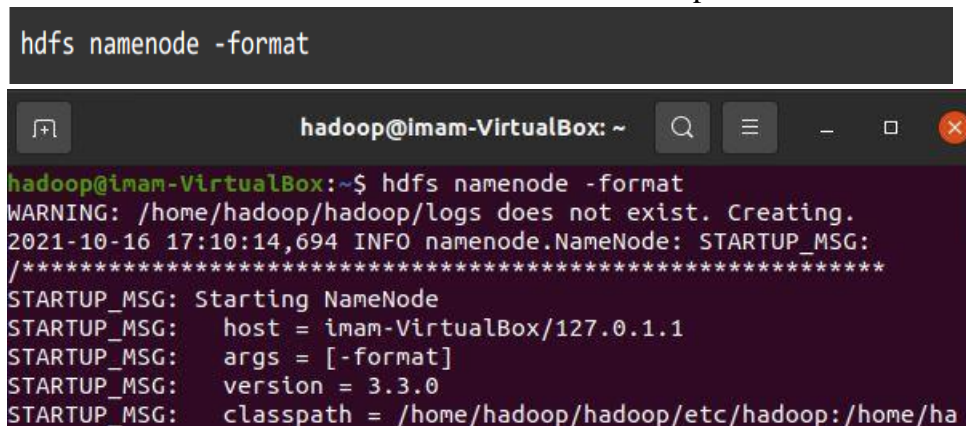
```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ nano $HADOOP_HOME/etc/hadoop/yarn-site.xml  
ml  
  
<configuration>  
  <property>  
    <name>yarn.nodemanager.aux-services</name>  
    <value>mapreduce_shuffle</value>  
  </property>  
</configuration>
```



```
hadoop@imam-VirtualBox: ~  
/home/hadoop/hadoop/etc/hadoop/yarn-site.xml Modified  
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  
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>
```

D. Menjalankan Hadoop

1. Tahap selanjutnya kita sudah bisa menjalankan hadoopnya dimulai dari perintah untuk format namenode terlebih dahulu di user hadoop.



```
hdfs namenode -format  
  
hadoop@imam-VirtualBox:~$ hdfs namenode -format  
WARNING: /home/hadoop/hadoop/logs does not exist. Creating.  
2021-10-16 17:10:14,694 INFO namenode.NameNode: STARTUP_MSG:  
/*****  
STARTUP_MSG: Starting NameNode  
STARTUP_MSG:  host = imam-VirtualBox/127.0.1.1  
STARTUP_MSG:  args = [-format]  
STARTUP_MSG:  version = 3.3.0  
STARTUP_MSG:  classpath = /home/hadoop/hadoop/etc/hadoop:/home/ha
```

2. Selanjutnya menjalankan hadoop cluster atau dfs (distributed file system) service dengan menghasilkan namenodes, datanodes, dan secondary namenodes.



```
start-dfs.sh
```

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ start-dfs.sh  
Starting namenodes on [localhost]  
pdsh@imam-VirtualBox: localhost: rcmd: socket: Permission denied  
Starting datanodes  
pdsh@imam-VirtualBox: localhost: rcmd: socket: Permission denied  
Starting secondary namenodes [imam-VirtualBox]  
pdsh@imam-VirtualBox: imam-VirtualBox: rcmd: socket: Permission denied  
hadoop@imam-VirtualBox:~$
```

Jika mengalami hal diatas lakukan perintah berikut untuk cek tipe rcmd.

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ pdsh -q -w localhost  
-- DSH-specific options --  
Separate stderr/stdout Yes  
Path prepended to cmd none  
Appended to cmd none  
Command: none  
Full program pathname /usr/bin/pdsh  
Remote program path /usr/bin/pdsh  
  
-- Generic options --  
Local username hadoop  
Local uid 1001  
Remote username hadoop  
Rcmd type rsh  
one ^C will kill pdsh No  
Connect timeout (secs) 10  
Command timeout (secs) 0  
Fanout 32  
Display hostname labels Yes  
Debugging No  
  
-- Target nodes --  
localhost  
hadoop@imam-VirtualBox:~$
```

Ubah menjadi ssh dengan perintah berikut

```
source ~/.bashrc
```

```
export PDSH_RCMD_TYPE=ssh
```

Kemudian jalankan perintah kembali

```
start-dfs.sh
```

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ start-dfs.sh  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [imam-VirtualBox]  
imam-VirtualBox: Warning: Permanently added 'imam-virtualbox' (ECDSA)  
to the list of known hosts.  
hadoop@imam-VirtualBox:~$
```

3. Selain itu jalankan juga YARN (Yet Another Resource Negotiator) service dengan menghasilkan resourcemanager dan nodemanagers.

```
start-yarn.sh
```

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ start-yarn.sh  
Starting resourcemanager  
resourcemanager is running as process 2465. Stop it first.  
Starting nodemanagers  
hadoop@imam-VirtualBox:~$
```

4. Setelah berhasil dijalankan, kita dapat melihat statusnya apakah sudah berhasil atau belum dengan jps (java virtual machine process status tool).

```
jps  
  
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ jps  
34321 NameNode  
2465 ResourceManager  
34658 SecondaryNameNode  
35043 NodeManager  
35172 Jps  
34463 DataNode  
hadoop@imam-VirtualBox:~$
```

5. Untuk menjalankan hadoop kita dapat menggunakan browser dengan IP 127.0.0.1 yang sudah kita setting sebelumnya di file core-site.xml dengan port 9870 dan 8088. Namun kita masih perlu menyesuaikan portnya agar bisa diakses dengan perintah berikut.

```
firewall-cmd --permanent --add-port=9870/tcp  
firewall-cmd --permanent --add-port=8088/tcp  
firewall-cmd --reload
```

Jika tidak berhasil dijalankan dan menampilkan hasil berikut, maka perlu melakukan instalasi firewalld terlebih dahulu menggunakan superuser.

```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ firewall-cmd --permanent --add-port=9870/tcp  
Command 'firewall-cmd' not found, but can be installed with:  
apt install firewallld  
Please ask your administrator.  
hadoop@imam-VirtualBox:~$
```



```
imam@imam-VirtualBox: /home/hadoop/hadoop
hadoop@imam-VirtualBox:~$ exit
logout
imam@imam-VirtualBox: /home/hadoop/hadoop$
```

```
exit
sudo apt install firewallld
```

```
imam@imam-VirtualBox: /home/hadoop/hadoop$ sudo apt install firewallld
[sudo] password for imam:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer
required:
  linux-headers-5.11.0-27-generic linux-hwe-5.11-headers-5.11.0-27
  linux-image-5.11.0-27-generic linux-modules-5.11.0-27-generic
  linux-modules-extra-5.11.0-27-generic
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  ipset libipset13 libnftables1 python3-decorator python3-firewall
python3-nftables python3-selinux python3-slip python3-slip-dbus
The following NEW packages will be installed:
  firewallld ipset libipset13 libnftables1 python3-decorator
python3-firewall python3-nftables python3-selinux python3-slip
python3-slip-dbus
0 upgraded, 10 newly installed, 0 to remove and 53 not upgraded.
Need to get 945 kB of archives.
After this operation, 5.382 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://id.archive.ubuntu.com/ubuntu focal/universe amd64 libnfta
bles1 amd64 0.9.3-2 [229 kB]
```

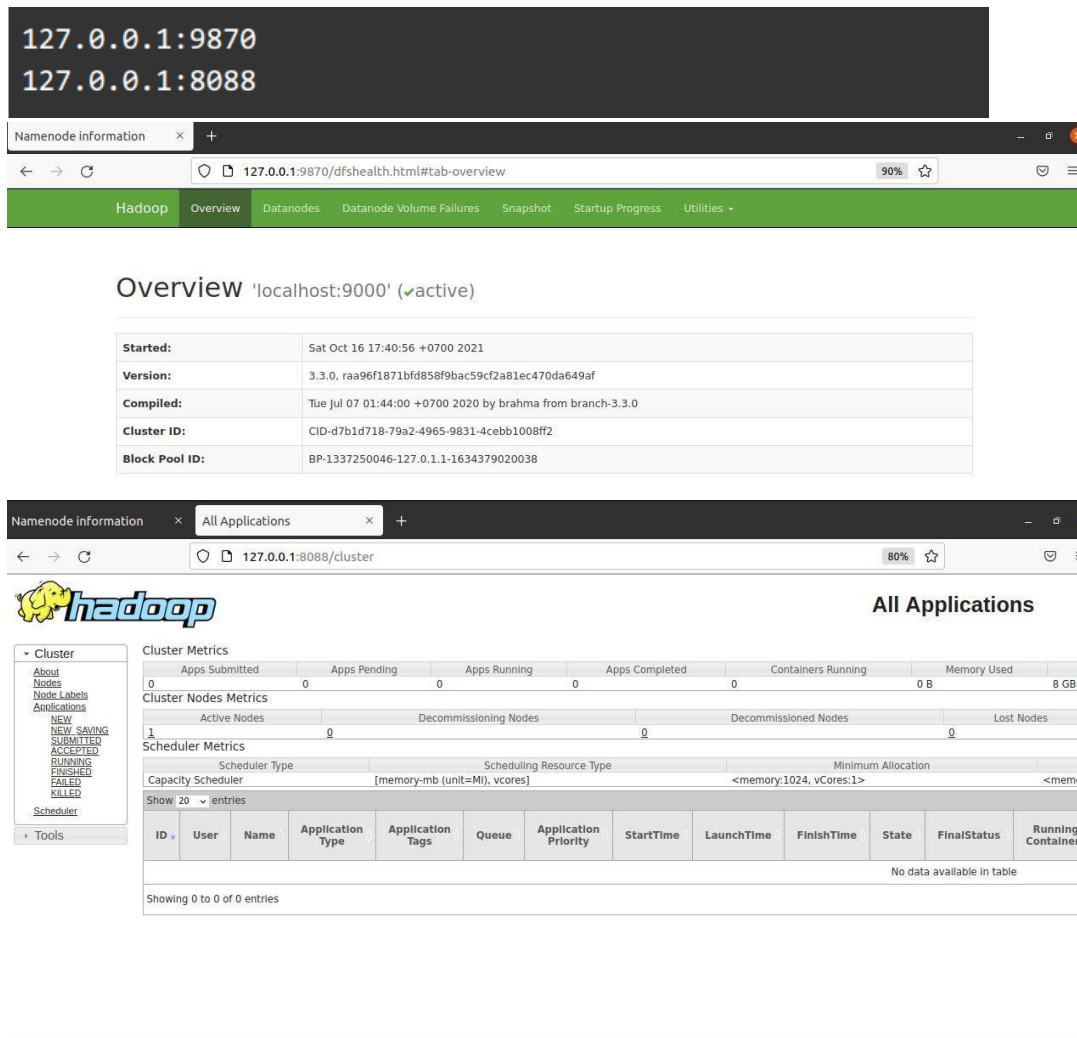
Setelah sudah berhasil diinstal maka coba kembali perintah di atas.

```
hadoop@imam-VirtualBox: ~
hadoop@imam-VirtualBox:~$ firewall-cmd --permanent --add-port=9870/tcp
success
hadoop@imam-VirtualBox:~$
```

```
hadoop@imam-VirtualBox: ~
hadoop@imam-VirtualBox:~$ firewall-cmd --permanent --add-port=8088/tcp
success
hadoop@imam-VirtualBox:~$
```

```
hadoop@imam-VirtualBox: ~
hadoop@imam-VirtualBox:~$ firewall-cmd --reload
success
hadoop@imam-VirtualBox:~$
```

6. Terakhir buka browser dan ketikkan URL berikut



The screenshot displays the Hadoop web interface. The top section shows the 'Overview' page for 'localhost:9000' (active). Below this, the 'All Applications' page is shown, featuring a sidebar with navigation links (Cluster, About, Nodes, Node Labels, Applications, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, Scheduler, Tools) and a main content area with various metrics tables.

Overview 'localhost:9000' (active)

Started:	Sat Oct 16 17:40:56 +0700 2021
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Tue Jul 07 01:44:00 +0700 2020 by brahma from branch-3.3.0
Cluster ID:	CID-d7b1d718-79a2-4965-9831-4cebb1008ff2
Block Pool ID:	BP-1337250046-127.0.1.1-1634379020038

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used
0	0	0	0	0 B	8 GB

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

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

Showing 0 to 0 of 0 entries

E. Testing Hadoop

1. Tahap terakhir kita akan coba melakukan testing bagaimana cara menggunakannya, kita buat terlebih dahulu contoh direktori.

```
hdfs dfs -mkdir /test1
hdfs dfs -mkdir /logs
```

```
hadoop@imam-VirtualBox:~$ hdfs dfs -mkdir /test1
hadoop@imam-VirtualBox:~$ hdfs dfs -mkdir /logs
hadoop@imam-VirtualBox:~$
```

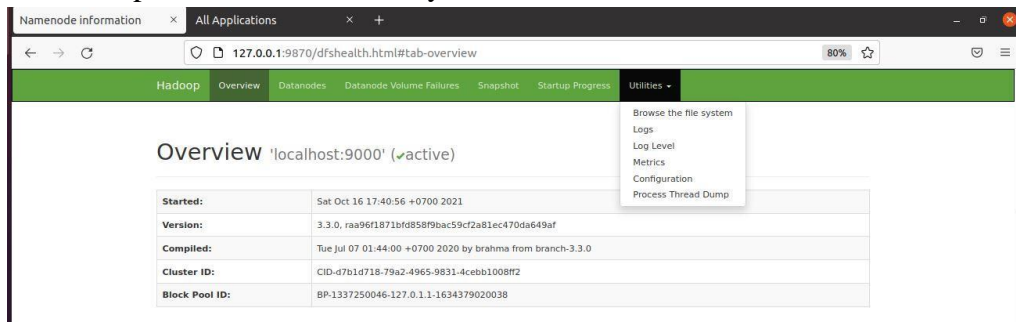
2. Perlu diingat direktori yang dibuat terletak di hdfs bukan di / (root) dan kita bisa memeriksanya dengan perintah berikut

```
hdfs dfs -ls /
```

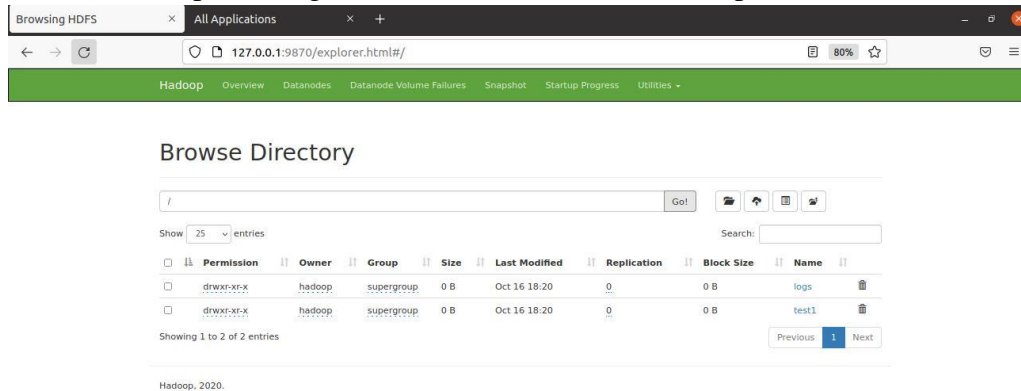


```
hadoop@imam-VirtualBox: ~$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x - hadoop supergroup          0 2021-10-16 18:20 /logs
drwxr-xr-x - hadoop supergroup          0 2021-10-16 18:20 /test1
hadoop@imam-VirtualBox:~$
```

Atau kalian bisa melihatnya langsung di browser dengan memilih menu Utilities kemudian pilih Browser the file system.



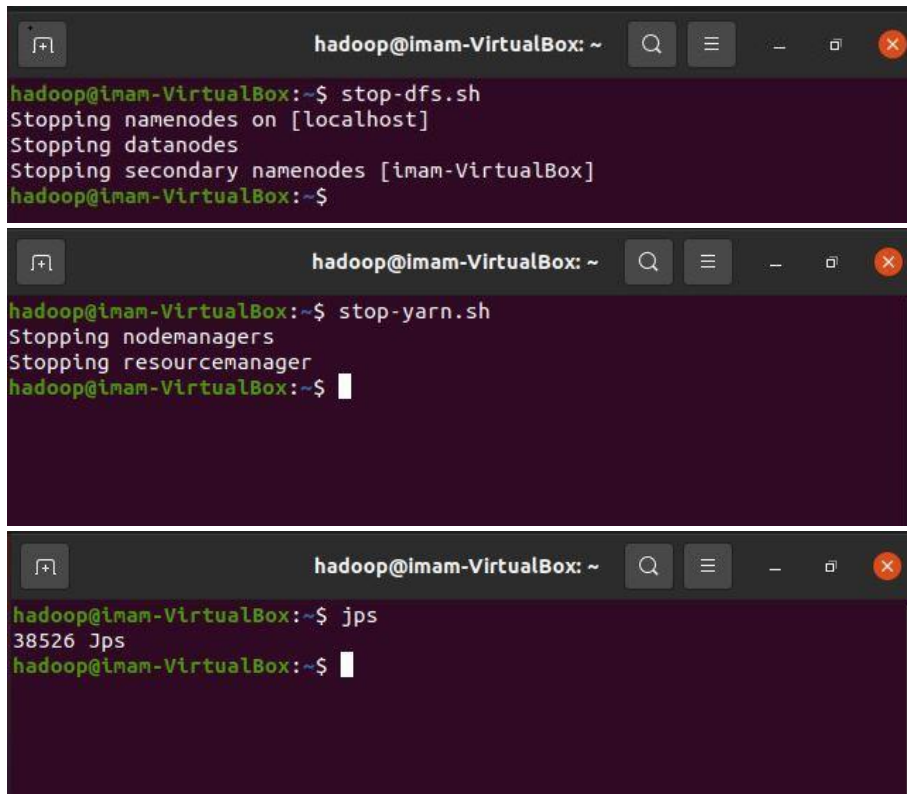
Maka akan terlihat 2 direktori file system sebelumnya dan untuk selanjutnya dapat kalian lakukan proses big data namun tidak akan dibahas pada modul ini.



F. Mematikan Hadoop

1. Jika kalian ingin mematikan service hadoopnya kalian dapat mengetikkan perintah berikut.

```
stop-dfs.sh
stop-yarn.sh
jps
```



```
hadoop@imam-VirtualBox: ~  
hadoop@imam-VirtualBox:~$ stop-dfs.sh  
Stopping namenodes on [localhost]  
Stopping datanodes  
Stopping secondary namenodes [imam-VirtualBox]  
hadoop@imam-VirtualBox:~$  
  
hadoop@imam-VirtualBox:~$ stop-yarn.sh  
Stopping nodemanagers  
Stopping resourcemanager  
hadoop@imam-VirtualBox:~$  
  
hadoop@imam-VirtualBox:~$ jps  
38526 Jps  
hadoop@imam-VirtualBox:~$
```

2. Jika kalian ingin menyalakan hadoop service lagi dapat juga dilakukan dengan perintah berikut, cara sebelumnya agar memastikan tidak ada yang error di setiap bagian.

```
start-all.sh  
stop-all.sh
```

Untuk Versi Windows, silakan pilih salah satu panduan ini :

- <https://www.teknologi-bigdata.com/2019/01/install-hadoop-2-standalone-windows.html>
- <https://medium.com/@pedro.a.hdez.a/hadoop-3-2-2-installation-guide-for-windows-10-454f5b5c22d3>
- <https://youtu.be/knAS0w-jiUk?si=uP17lZSsgKsmS7zg>

Referensi

- <https://medium.com/skyshidigital/teknologi-big-data-dengan-hadoop-d8a2e93791a8>
- <https://tecadmin.net/install-hadoop-on-ubuntu-20-04/>
- Modul Praktikum 1: Instalasi Hadoop, Mata Kuliah Big Data 20202