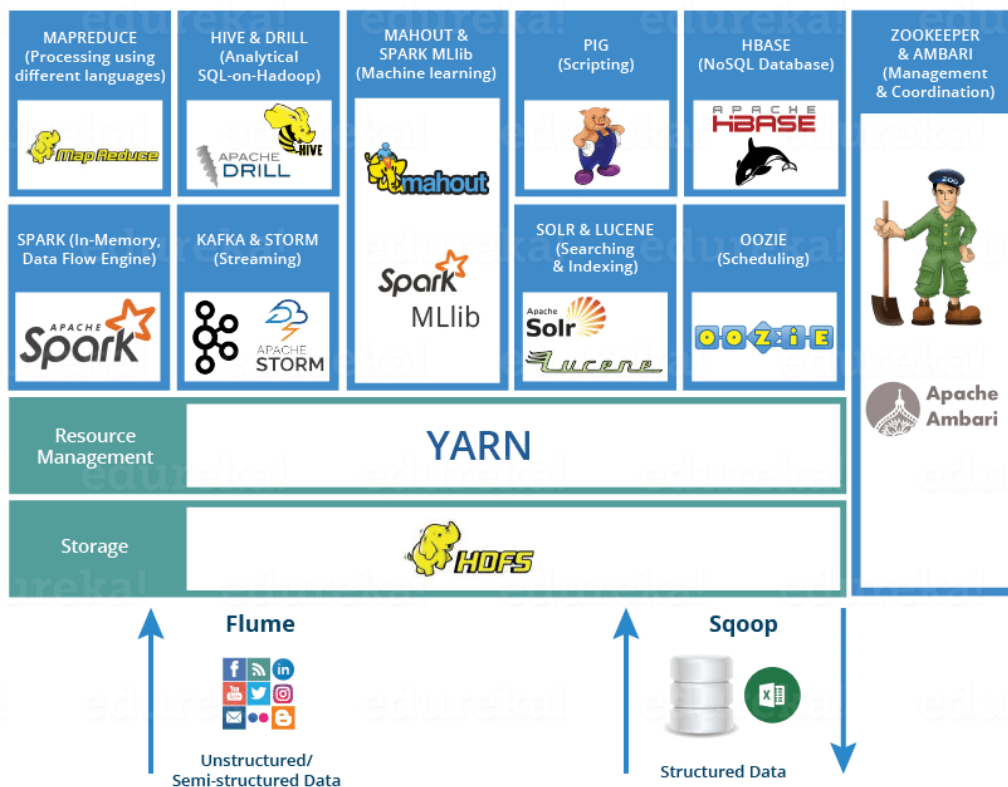


## Modul 7

### Instalasi dan Konfigurasi Hadoop

#### RINGKASAN MATERI:

**Hadoop** merupakan *software framework open source* yang memungkinkan pemrosesan data berukuran besar (big data) secara terdistribusi dengan melibatkan berkluster-kluster komputer. Hadoop didesain untuk dapat bekerja secara efektif baik dalam skala terkecil yang hanya melibatkan satu server hingga skala besar yang mempekerjakan ribuan komputer dimana masing-masing komputer tersebut memfasilitasi komputasi dan penyimpanan data secara lokal. Untuk menjamin *High Availability*, Hadoop tidak menggantungkannya pada *hardware* yang digunakan, tetapi *framework* Hadoop itu sendiri telah didesain untuk dapat mendeteksi dan menangani *failure* pada level/layer aplikasi. Hadoop sendiri berbasis Java yang berada di bawah lisensi Apache.

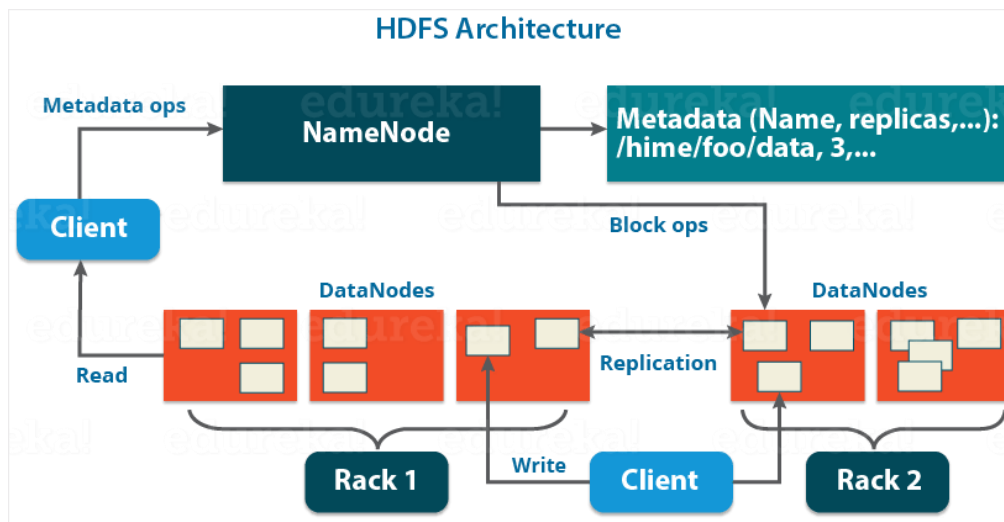


Gambar Ekosistem Hadoop Pada Umumnya

Dalam Hadoop, terdapat istilah **Hadoop Distributed File System (HDFS)**. HDFS merupakan *open source project* yang dikembangkan oleh Apache Software Foundation dan merupakan subproject dari Apache Hadoop. Apache mengembangkan HDFS berdasarkan konsep dari Google File System (GFS) dan oleh karenanya sangat mirip dengan GFS baik ditinjau dari konsep logika, struktur fisik, maupun cara kerjanya. Sebagai layer penyimpanan data di Hadoop, HDFS adalah sebuah sistem *file* berbasis Java yang *fault-tolerant*, terdistribusi, dan *scalable*. HDFS dirancang agar dapat diaplikasikan pada kluster dan dapat dijalankan dengan menggunakan *proprietary* atau *commodity server*. HDFS ini pada dasarnya adalah sebuah

direktori dimana data disimpan yang bekerja sesuai dengan spesifikasi dari Hadoop. Data tersimpan dalam *cluster* yang terdiri dari banyak node komputer/server yang masing-masing sudah terinstal Hadoop.

HDFS memiliki komponen-komponen utama berupa NameNode dan DataNode. NameNode adalah sebuah komputer yang bertindak sebagai master, sedangkan DataNode adalah komputer-komputer dalam Hadoop Cluster yang bertugas sebagai *slaves* atau anak buah. NameNode bertanggung jawab menyimpan informasi tentang penempatan block data dalam Hadoop Cluster. Selain itu NameNode bertanggung jawab mengorganisir dan mengontrol block data yang disimpan tersebar dalam komputer-komputer yang menyusun Hadoop Cluster. Sedangkan DataNode bertugas menyimpan block-block data yang dialamatkan kepadanya, dan secara berkala melaporkan kondisinya kepada NameNode.



Gambar Arsitektur HDFS Hadoop

## LANGKAH-LANGKAH PERSIAPAN:

(Disclaimer)

- JANGAN LUPA untuk membuat snapshot di beberapa kondisi, khususnya ketika melakukan instalasi atau hal-hal yang sekiranya dirasa penting bagi praktikan agar jika terjadi error dapat melakukan *recovery* dan *error handling* dengan mudah.
- Perhatikan PATH serta PENAMAAN masing-masing file, sesuaikan dengan yang sudah diterima.

Referensi :

- <https://pinetools.com/random-file-generator>
- <https://phoenixnap.com/kb/install-hadoop-ubuntu>
- <http://malifauzi.lecture.ub.ac.id/2019/02/instalasi-hadoop-pada-ubuntu-server-ubuntu-desktop/>
- <http://hadoop.apache.org/>
- <https://www.coursera.org/learn/big-data-introduction/supplement/KWZQG/copy-your-data-into-the-hadoop-distributed-file-system-hdfs-instructions>

## A. Mengecek Versi Java

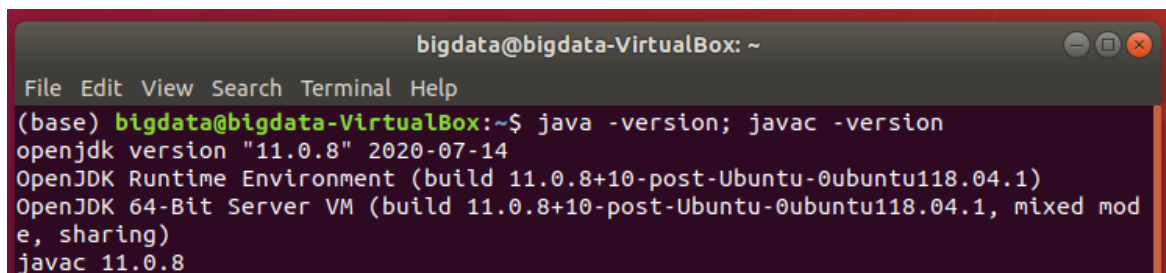
1. Mengecek versi java. Dalam modul ini menggunakan versi 11.0.8. Disarankan untuk menyamakan versi, namun juga tidak masalah jika tidak sama (setidaknya mendekati versi 11.0 ke atas). Gunakan *command* berikut.

→ `javac --version`

atau

→ `java -version`

Output berupa informasi versi Java yang dimiliki, kurang lebih serupa dengan gambar di bawah ini.

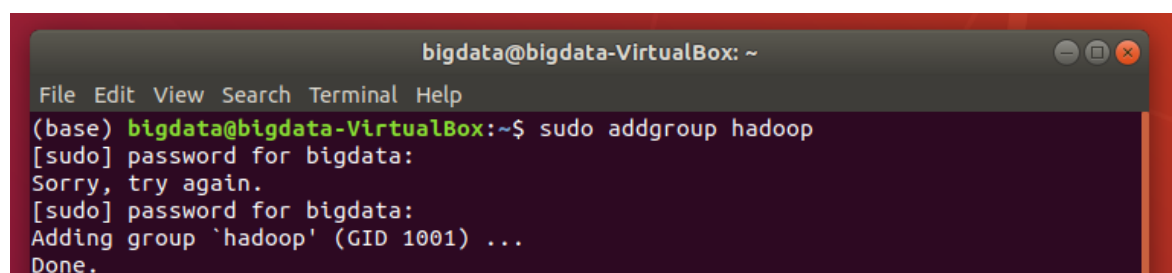
A terminal window titled 'bigdata@bigdata-VirtualBox: ~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '(base) bigdata@bigdata-VirtualBox:~\$ java -version; javac -version' has been executed. The output is: 'openjdk version "11.0.8" 2020-07-14', 'OpenJDK Runtime Environment (build 11.0.8+10-post-Ubuntu-0ubuntu118.04.1)', 'OpenJDK 64-Bit Server VM (build 11.0.8+10-post-Ubuntu-0ubuntu118.04.1, mixed mode, sharing)', and 'javac 11.0.8'.

```
bigdata@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
(base) bigdata@bigdata-VirtualBox:~$ java -version; javac -version
openjdk version "11.0.8" 2020-07-14
OpenJDK Runtime Environment (build 11.0.8+10-post-Ubuntu-0ubuntu118.04.1)
OpenJDK 64-Bit Server VM (build 11.0.8+10-post-Ubuntu-0ubuntu118.04.1, mixed mode, sharing)
javac 11.0.8
```

## B. Membuat Group dan User

1. Menambahkan group baru dengan nama 'hadoop' menggunakan command di bawah ini. Jika berhasil akan memunculkan informasi 'Done':

`sudo addgroup hadoop`

A terminal window titled 'bigdata@bigdata-VirtualBox: ~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '(base) bigdata@bigdata-VirtualBox:~\$ sudo addgroup hadoop' has been executed. The output is: '[sudo] password for bigdata:', 'Sorry, try again.', '[sudo] password for bigdata:', 'Adding group \'hadoop\' (GID 1001) ...', and 'Done.'. Note that the password was entered incorrectly the first time.

```
bigdata@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
(base) bigdata@bigdata-VirtualBox:~$ sudo addgroup hadoop
[sudo] password for bigdata:
Sorry, try again.
[sudo] password for bigdata:
Adding group 'hadoop' (GID 1001) ...
Done.
```

2. Membuat user baru dalam group hadoop menggunakan command di bawah ini. Hasil seharusnya menyerupai gambar di bawah ini.

`sudo adduser -ingroup hadoop hduser`

## PERHATIKAN!

- UNIX password **gunakan password yang mudah diingat**
- **Fullname: hduser** (samakan, untuk kemudahan praktikum)
- Room Number, Work Phone, Home Phone, Other: default (langsung enter saja)
- Untuk konfirmasi, tekan Y. Dapat menekan n jika belum yakin dan ingin mengulang prosesnya.

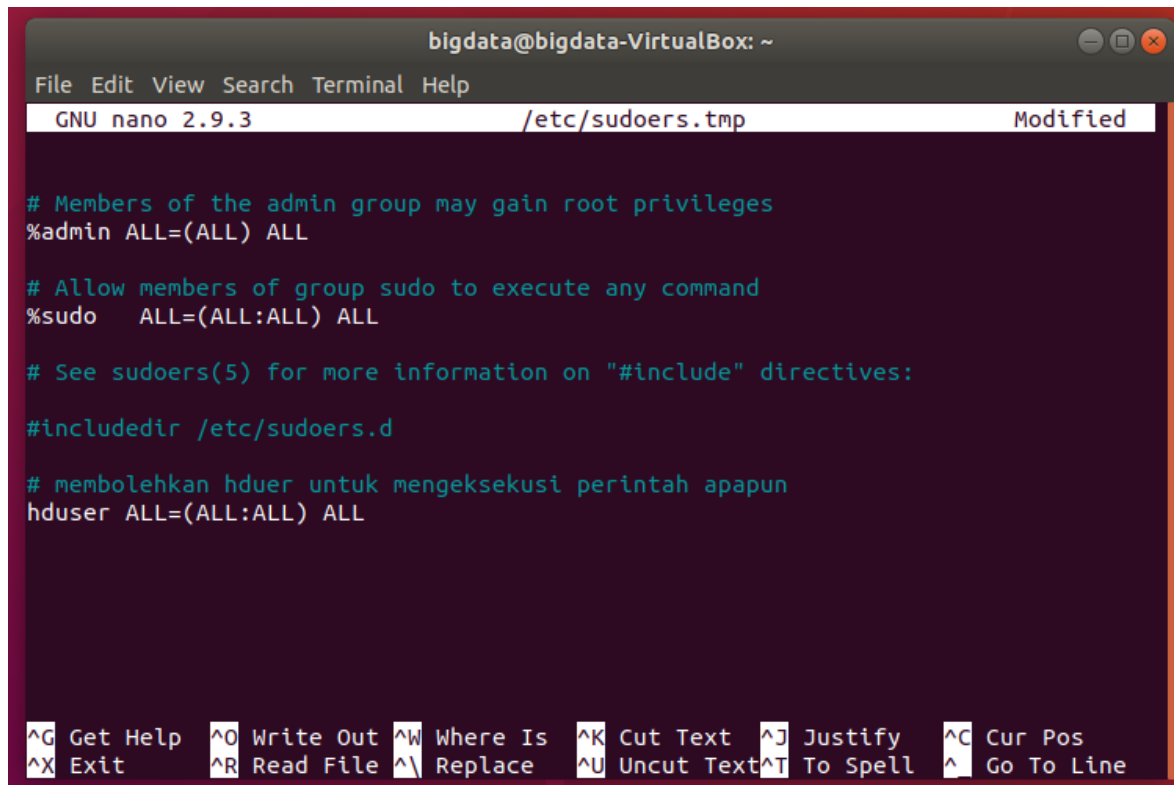
```
(base) bigdata@bigdata-VirtualBox:~$ sudo adduser -ingroup hadoop hduser
Adding user `hduser' ...
Adding new user `hduser' (1001) with group `hadoop' ...
Creating home directory `/home/hduser' ...
Copying files from `/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password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

3. Edit visudo untuk user hduser dengan command  
sudo visudo

Kemudian menambahkan baris

```
hduser ALL=(ALL:ALL) ALL
```

Sesuaikan dengan gambar di bawah ini.



```
bigdata@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 /etc/sudoers.tmp Modified

# Members of the admin group may gain root privileges
%admin ALL=(ALL) ALL

# Allow members of group sudo to execute any command
%sudo  ALL=(ALL:ALL) ALL

# See sudoers(5) for more information on "#include" directives:

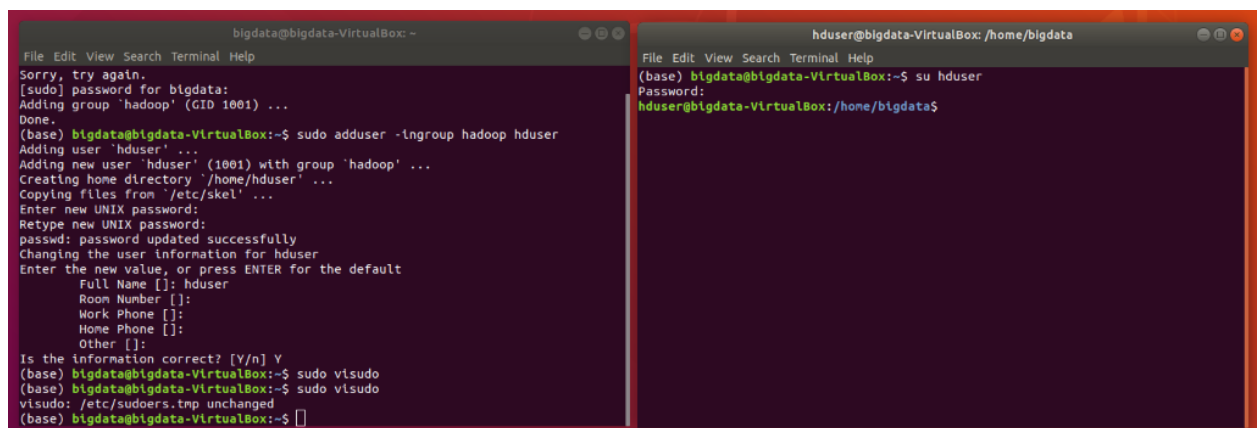
#includedir /etc/sudoers.d

# membolehkan hduer untuk mengeksekusi perintah apapun
hduser ALL=(ALL:ALL) ALL

^G Get Help  ^O Write Out ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File ^\ Replace   ^U Uncut Text ^T To Spell  ^_ Go To Line
```

Jika sudah, dapat keluar dari editor dengan menekan Ctrl + X → klik Y → tekan enter.

4. Buka terminal baru dan masuk sebagai user hduser (lihat terminal yang kanan). Masukkan password sesuai yang telah diinput pada langkah nomor dua.  
su hduser



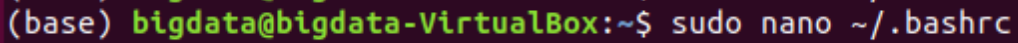
```
bigdata@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
Sorry, try again.
[sudo] password for bigdata:
Adding group 'hadoop' (GID 1001) ...
Done.
(base) bigdata@bigdata-VirtualBox:~$ sudo adduser -ingroup hadoop hduser
Adding user 'hduser' ...
Adding new user 'hduser' (1001) with group 'hadoop' ...
Creating home directory '/home/hduser' ...
Copying files from '/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
  Full Name []: hduser
    Room Number []:
    Work Phone []:
    Home Phone []:
    Other []:
Is the information correct? [Y/n] Y
(base) bigdata@bigdata-VirtualBox:~$ sudo visudo
(base) bigdata@bigdata-VirtualBox:~$ sudo visudo
visudo: /etc/sudoers.tmp unchanged
(base) bigdata@bigdata-VirtualBox:~$

hduser@bigdata-VirtualBox: /home/bigdata
File Edit View Search Terminal Help
(base) bigdata@bigdata-VirtualBox:~$ su hduser
Password:
hduser@bigdata-VirtualBox: /home/bigdata$
```

5. **DISARANKAN** melakukan **SNAPSHOT** ketika sudah sampai pada langkah ini untuk mempermudah mengatasi *error* yang kemungkinan akan terjadi kedepannya.

## C. Setting Java Environment

1. Kembali gunakan **terminal root** dan ubah `.bash` dengan command berikut.  
`sudo nano ~/.bashrc`

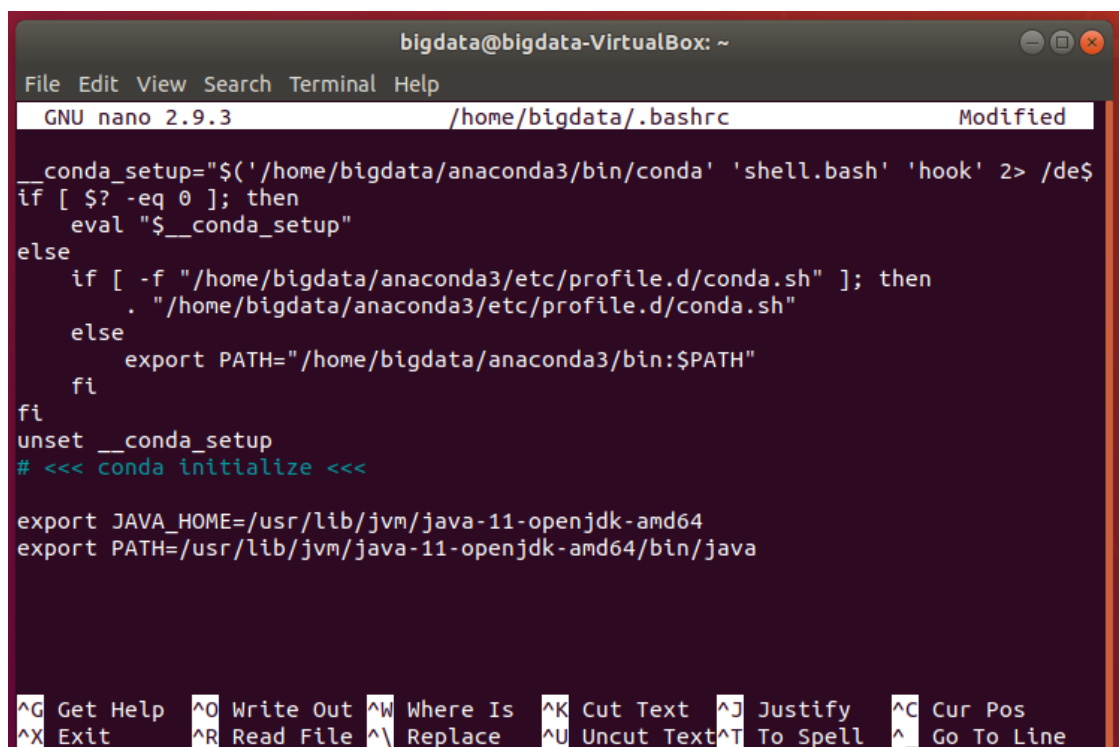


```
(base) bigdata@bigdata-VirtualBox:~$ sudo nano ~/.bashrc
```

2. Tambahkan `JAVA_HOME` dan `PATH` dengan mengetikkan berikut ini sesuai dengan yang terdapat pada gambar di bawah ini.

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

```
export PATH=/usr/lib/jvm/java-11-openjdk-amd64/bin/java
```



```
bigdata@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 /home/bigdata/.bashrc Modified

__conda_setup="$('/home/bigdata/anaconda3/bin/conda' 'shell.bash' 'hook' 2> /de$
if [ $? -eq 0 ]; then
    eval "$__conda_setup"
else
    if [ -f "/home/bigdata/anaconda3/etc/profile.d/conda.sh" ]; then
        . "/home/bigdata/anaconda3/etc/profile.d/conda.sh"
    else
        export PATH="/home/bigdata/anaconda3/bin:$PATH"
    fi
fi
unset __conda_setup
# <<< conda initialize <<<

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export PATH=/usr/lib/jvm/java-11-openjdk-amd64/bin/java

^G Get Help  ^O Write Out ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File ^\ Replace  ^U Uncut Text ^T To Spell  ^_ Go To Line
```

Jika sudah, dapat keluar dari editor dengan menekan `Ctrl + X` → klik `Y` → tekan enter.

3. Update environment dengan command `source ~/.bashrc`
4. Mengecek apakah sudah masuk pada environment dengan command berikut.  
`printenv`

```
(base) bigdata@bigdata-VirtualBox:~$ printenv
CLUTTER_IM_MODULE=xim
CONDA_SHLVL=1
LS_COLORS=rs=0:di=01;34:ln=01;36:mh=00:pi=40;33:so=01;35:do=01;35:bd=40;33;01:cd
=40;33;01:or=40;31;01:mi=00:su=37;41:sg=30;43:ca=30;41:tw=30;42:ow=34;42:st=37;4
4:ex=01;32:*.tar=01;31:*.tgz=01;31:*.arc=01;31:*.arj=01;31:*.taz=01;31:*.lha=01;
31:*.lzh=01;31:*.lzh=01;31:*.lzm=01;31:*.tlz=01;31:*.txz=01;31:*.tzo=01;31:*.t7
...

```

Seharusnya terdapat bagian ini jika sudah masuk dalam environment-nya.

```
conda activate /usr/lib/jvm/java-11-openjdk-amd64
JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
XDG_VTNR=2
SSH_AUTH_SOCK=/run/user/1000/keyring/ssh
S_COLORS=auto

```

Apabila *command* **printenv**, **sudo**, dan **su** tidak dikenali oleh terminal seperti pada gambar di bawah ini, maka jalankan *command* `export PATH=/usr/bin:/bin` atau `export PATH="/usr/bin:$PATH"`

```
(base) nurlaita@nurlaita-VirtualBox:~$ su hduser
Command 'su' is available in '/bin/su'
The command could not be located because '/bin' is not included in the PATH envi
ronment variable.
su: command not found
(base) nurlaita@nurlaita-VirtualBox:~$ export PATH=/usr/bin:/bin
(base) nurlaita@nurlaita-VirtualBox:~$ su hduser
Password:

```

Namun *command* ini bersifat sementara (*temporary*). Perlu diperhatikan ketika melakukan perubahan user dan perubahan *state* terminal.

5. Membuat jdk sebagai default JVM menggunakan kedua *command* berikut
  - `sudo update-alternatives --install /usr/bin/java java /usr/lib/jvm/java-11-openjdk-amd64/bin/java 1`
  - `sudo update-alternatives --install /usr/bin/javac javac /usr/lib/jvm/java-11-openjdk-amd64/bin/javac 1`

```
(base) bigdata@bigdata-VirtualBox:~$ sudo update-alternatives --install /usr/bin
/java java /usr/lib/jvm/java-11-openjdk-amd64/bin/java 1
update-alternatives: warning: forcing reinstallation of alternative /usr/lib/jvm
/java-11-openjdk-amd64/bin/java because link group java is broken
(base) bigdata@bigdata-VirtualBox:~$ sudo update-alternatives --install /usr/bin
/javac javac /usr/lib/jvm/java-11-openjdk-amd64/bin/javac 1
update-alternatives: warning: forcing reinstallation of alternative /usr/lib/jvm
/java-11-openjdk-amd64/bin/javac because link group javac is broken

```

6. Melakukan verifikasi update java menggunakan *command* berikut.
  - `sudo update-alternatives --config java`



→ `sudo update-alternatives --config javac`

```
(base) bigdata@bigdata-VirtualBox:~$ sudo update-alternatives --config java
There is only one alternative in link group java (providing /usr/bin/java): /usr
/lib/jvm/java-11-openjdk-amd64/bin/java
Nothing to configure.
(base) bigdata@bigdata-VirtualBox:~$ sudo update-alternatives --config javac
There is only one alternative in link group javac (providing /usr/bin/javac): /u
sr/lib/jvm/java-11-openjdk-amd64/bin/javac
Nothing to configure.
```

7. Mengecek kembali versi java menggunakan command berikut.

`java -version`

```
(base) bigdata@bigdata-VirtualBox:~$ java -version
openjdk version "11.0.8" 2020-07-14
OpenJDK Runtime Environment (build 11.0.8+10-post-Ubuntu-0ubuntu118.04.1)
OpenJDK 64-Bit Server VM (build 11.0.8+10-post-Ubuntu-0ubuntu118.04.1, mixed mod
e, sharing)
```

8. Mengecek kembali lokasi java menggunakan command berikut.

`which java`

```
hduser@bigdata-VirtualBox:/home/bigdata$ which java
/usr/bin/java
hduser@bigdata-VirtualBox:/home/bigdata$
```

## D. Setting Open SSH

1. Install OpenSSH server dan client menggunakan command berikut.

`sudo apt install openssh-server openssh-client -y`

Seharusnya hasil akan keluar seperti di bawah ini.

```
(base) bigdata@bigdata-VirtualBox:~$ sudo apt install openssh-server openssh-cl
ient -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:7.6p1-4ubuntu0.3).
openssh-client set to manually installed.
The following packages were automatically installed and are no longer required:
  linux-headers-5.4.0-42-generic linux-hwe-5.4-headers-5.4.0-42
  linux-image-5.4.0-42-generic linux-modules-5.4.0-42-generic
  linux-modules-extra-5.4.0-42-generic
```

Jika terdapat error, dapat menggunakan command berikut..



```
sudo apt install openssh-server openssh-client
```

```
(base) nurlaita@nurlaita-VirtualBox:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
E: Unable to locate package -y

(base) nurlaita@nurlaita-VirtualBox:~$ sudo apt install openssh-server openssh-client
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:7.6p1-4ubuntu0.3).
openssh-server is already the newest version (1:7.6p1-4ubuntu0.3).
The following packages were automatically installed and are no longer required:
  linux-hwe-5.4-headers-5.4.0-42 linux-hwe-5.4-headers-5.4.0-48
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  ncurses-term openssh-sftp-server ssh-import-id
Suggested packages:
  molly-guard monkeysphere rssh 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 21 not upgraded.
Need to get 637 kB of archives.
```

2. Memverifikasi instalasi dengan proses yang sama untuk mengecek apakah sudah terpasang atau belum.

```
sudo apt install openssh-server openssh-client -y
```

```
fawwaz@fawwaz-VirtualBox:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:7.6p1-4ubuntu0.3).
openssh-server is already the newest version (1:7.6p1-4ubuntu0.3).
The following packages were automatically installed and are no longer required:
  linux-hwe-5.4-headers-5.4.0-42 linux-hwe-5.4-headers-5.4.0-48
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 32 not upgraded.
```

Jika terdapat error, dapat menggunakan command berikut.

```
sudo apt install openssh-server openssh-client
```

3. Pindah ke user hduser, atau ke terminal yang sebelumnya sudah dibuka dengan user hduser. Kemudian Generate SSH key pair dan definisikan lokasinya untuk dapat disimpan menggunakan command berikut.

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

```
hduser@bigdata-VirtualBox: /home/bigdata
File Edit View Search Terminal Help
hduser@bigdata-VirtualBox:/home/bigdata$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hduser/.ssh'.
Your identification has been saved in /home/hduser/.ssh/id_rsa.
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:Gz0kTCMqUbccpLDW32F6k9AY08fhpeBhFpvDnyxigEw hduser@bigdata-VirtualBox
The key's randomart image is:
+---[RSA 2048]---+
| E...B.Oo...|
| o * = /.Boo|
| * = = %.+ |
| . . o = X .|
| = S * |
| . o = . |
| . |
+-----[SHA256]-----+
```

4. Gunakan command berikut untuk menyimpan public key sebagai `authorized_keys` dalam direktori `ssh`.

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

```
hduser@bigdata-VirtualBox:/home/bigdata$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hduser@bigdata-VirtualBox:/home/bigdata$
```

5. Set hak akses untuk user menggunakan command berikut.

```
chmod 0600 ~/.ssh/authorized_keys
```

```
hduser@fawwaz-VirtualBox:/home/fawwaz$ chmod 0600 ~/.ssh/authorized_keys
hduser@fawwaz-VirtualBox:/home/fawwaz$
```

6. Seharusnya sekarang user telah bisa SSH tanpa harus memasukkan password setiap kali mencobanya. Perlu untuk memverifikasi dengan *set up user* `hduser` pada SSH untuk `localhost` menggunakan command berikut.

```
ssh localhost
```

Kemudian klik `yes`. Maka user `hduser` sudah dapat melakukan koneksi dengan `localhost` tanpa password.

```
hduser@bigdata-VirtualBox:/home/bigdata$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:PxE4YflULu3xLqBMrhdX1TjmHWGJJJoAbFhNc3BSZh5k.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.5 LTS (GNU/Linux 5.4.0-48-generic x86_64)
```

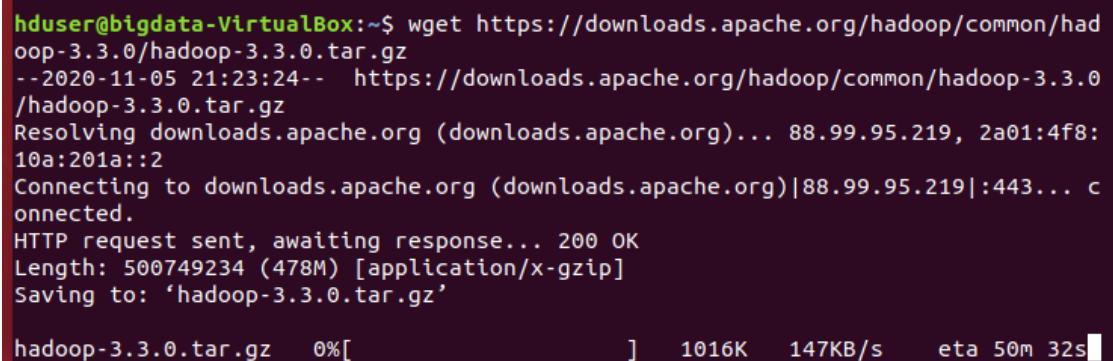
(opsional) !!! Jika masih meminta password, dapat mengetik command berikut.

```
sudo apt-get update
```

## E. Install Hadoop

1. Gunakan user hduser kemudian download **hadoop package** menggunakan command berikut. Besar file kurang lebih 478M sehingga membutuhkan waktu untuk menunggu proses selesai.

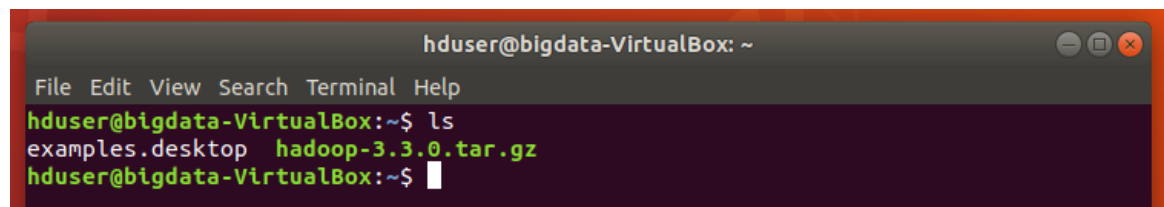
```
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
```



```
hduser@bigdata-VirtualBox:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
--2020-11-05 21:23:24-- 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, 2a01:4f8:10a:201a::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%[                  ] 1016K  147KB/s  eta 50m 32s
```

Jika sudah, kemudian dapat melakukan pengecekan apakah sudah ada filenya dengan command ls. Seharusnya hasilnya seperti di bawah ini.



```
hduser@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
hduser@bigdata-VirtualBox:~$ ls
examples.desktop  hadoop-3.3.0.tar.gz
hduser@bigdata-VirtualBox:~$
```

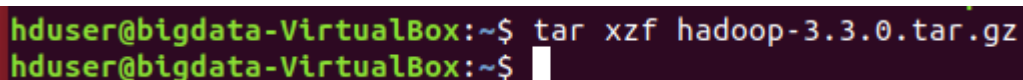
(opsional) !!! Jika mendownload secara terpisah, maka pastikan owner dari berkas adalah hduser dengan perintah

```
sudo chown hduser hadoop-3.3.0.tar.gz
```

2. Ekstrak hasil download menggunakan command berikut.

```
tar xzf hadoop-3.3.0.tar.gz
```

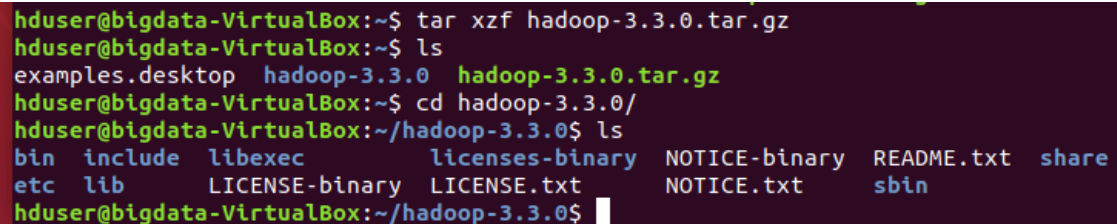
Tunggu hingga proses ekstraksi selesai.



```
hduser@bigdata-VirtualBox:~$ tar xzf hadoop-3.3.0.tar.gz
hduser@bigdata-VirtualBox:~$
```

3. Cek lokasi hasil ekstrak menggunakan command berikut.  
`Cd hadoop-3.3.0/`

Kemudian command `ls`.



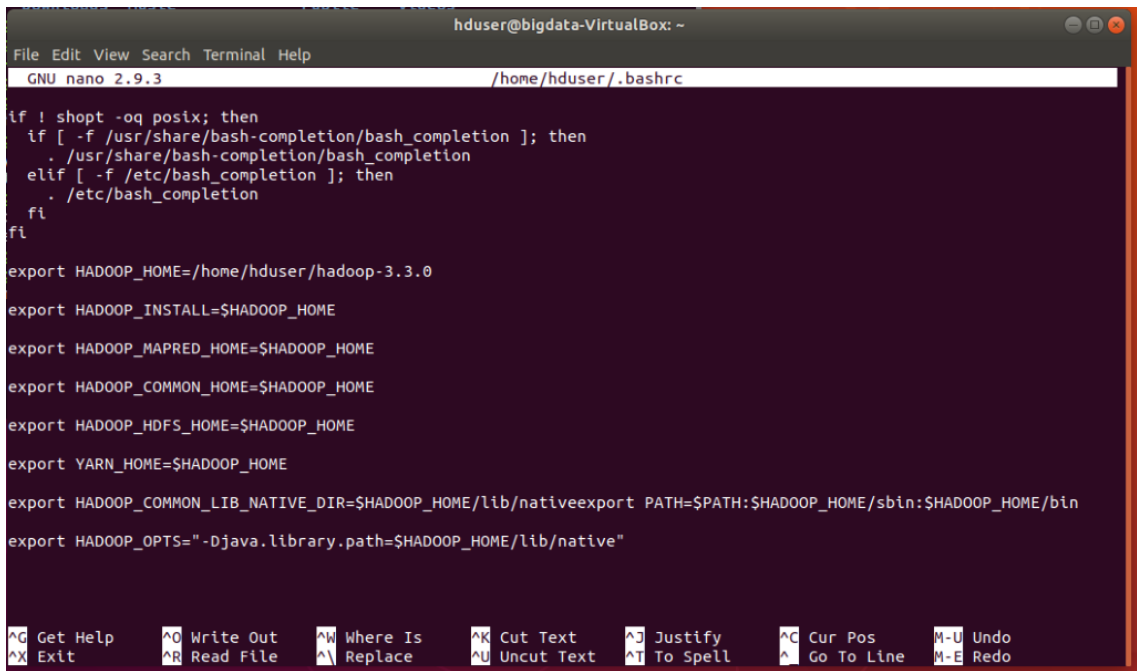
```
hduser@bigdata-VirtualBox:~$ tar xzf hadoop-3.3.0.tar.gz
hduser@bigdata-VirtualBox:~$ ls
examples.desktop  hadoop-3.3.0  hadoop-3.3.0.tar.gz
hduser@bigdata-VirtualBox:~$ cd hadoop-3.3.0/
hduser@bigdata-VirtualBox:~/hadoop-3.3.0$ ls
bin  include  libexec      licenses-binary  NOTICE-binary  README.txt  share
etc  lib       LICENSE-binary  LICENSE.txt      NOTICE.txt     sbin
```

4. **DISARANKAN** melakukan **SNAPSHOT** ketika sudah sampai pada langkah ini untuk mempermudah mengatasi *error* yang kemungkinan akan terjadi kedepannya.

## F. Setting Hadoop dengan Single Node / Psedo Distributed Mode

1. Setting variable pada `bashrc` dengan user `hduser` `.bashrc` yang dapat diakses menggunakan command `sudo nano ~/.bashrc`  
Jika diminta, Masukkan password, maka akan terbuka editornya. Kemudian tambahkan baris setting tersebut dengan menambahkan berikut ini:

```
export HADOOP_HOME=/home/hduser/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"
```

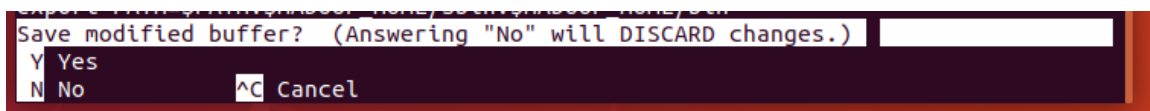


```
GNU nano 2.9.3 /home/hduser/.bashrc

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

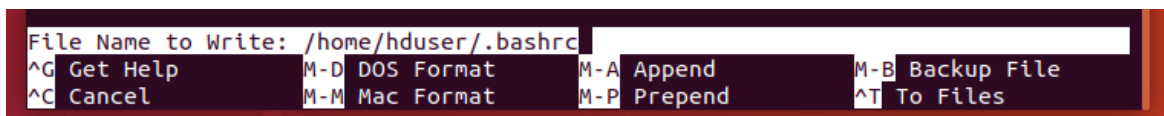
export HADOOP_HOME=/home/hduser/hadoop-3.3.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/nativeexport PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin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

Kemudian control x untuk keluar, Save dengan tuliskan Y



```
Save modified buffer? (Answering "No" will DISCARD changes.)
Y Yes
N No
^C Cancel
```

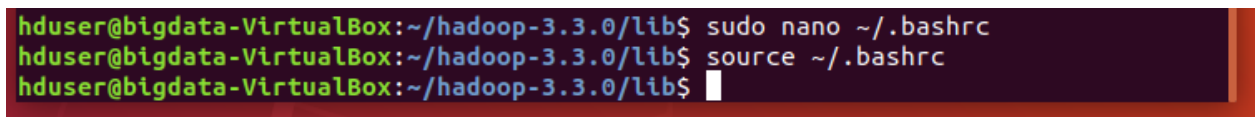
untuk file name biarkan sama. Langsung enter saja.



```
File Name to Write: /home/hduser/.bashrc
^G Get Help      M-D DOS Format   M-A Append      M-B Backup File
^C Cancel        M-M Mac Format   M-P Prepend     ^T To Files
```

2. Jalankan perintah untuk menyimpan setting untuk envirotnmen yang berjalan dengan menggunakan perintah :

`source ~/.bashrc`

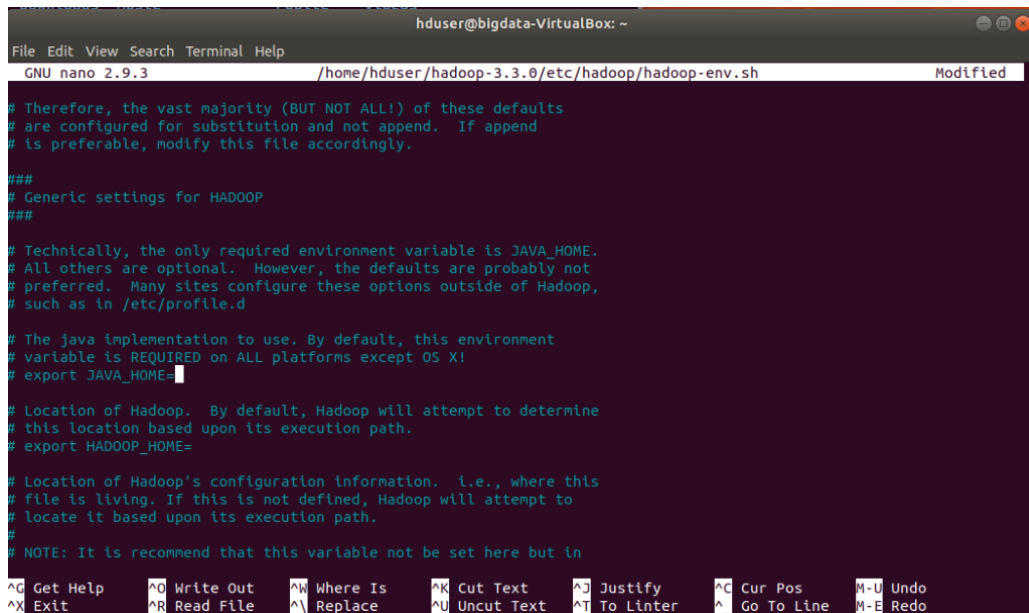


```
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/lib$ sudo nano ~/.bashrc
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/lib$ source ~/.bashrc
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/lib$
```

Jika terdapat error, cek lagi perhuruf untuk setting hadoopnya pada langkah 1.

3. Pengeditan berkas `hadoop-env.sh` dengan command:

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



```
hduser@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 /home/hduser/hadoop-3.3.0/etc/hadoop/hadoop-env.sh Modified

# Therefore, the vast majority (BUT NOT ALL!) of these defaults
# are configured for substitution and not append. If append
# is preferable, modify this file accordingly.

###
# Generic settings for HADOOP
###

# Technically, the only required environment variable is JAVA_HOME.
# All others are optional. However, the defaults are probably not
# 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=

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

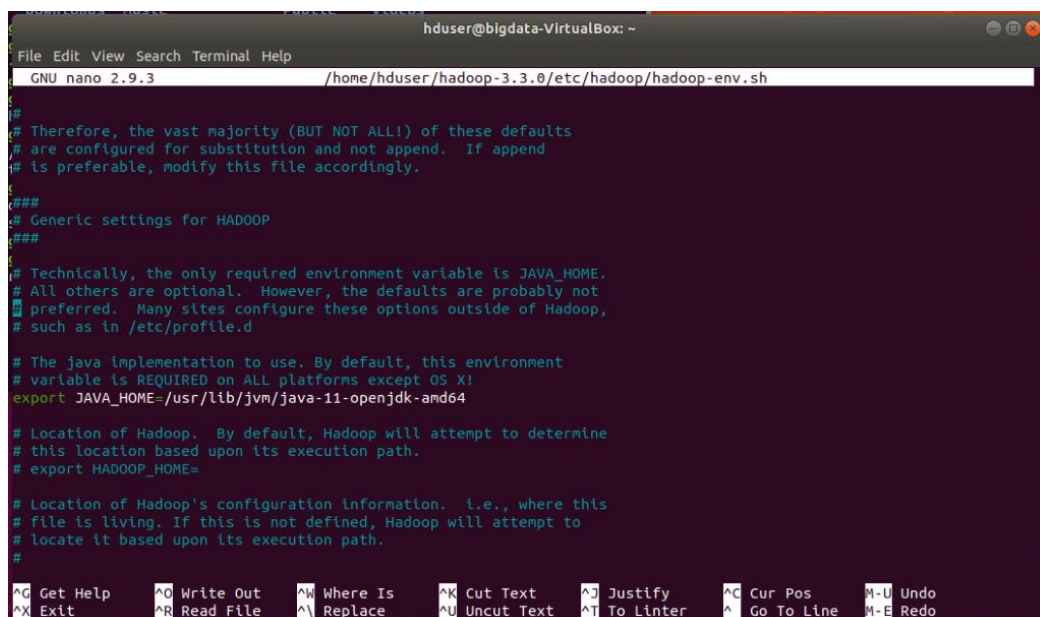
# Location of Hadoop's configuration information. i.e., where this
# file is living. If this is not defined, Hadoop will attempt to
# locate it based upon its execution path.
#
# NOTE: It is recommend that this variable not be set here but in

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos ^M-U Undo
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Linter ^_ Go To Line ^-E Redo
```

Cari JAVA\_HOME seperti pada gambar diatas, bisa dicari dengan Ctrl + W

Kemudian hapus # dan ganti alamatnya sesuai dengan path sesuai dengan langkah C2.  
Path yang dimasukkan tidak sampai bin seperti pada gambar dibawah :

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```



```
hduser@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 /home/hduser/hadoop-3.3.0/etc/hadoop/hadoop-env.sh

#
# Therefore, the vast majority (BUT NOT ALL!) of these defaults
# are configured for substitution and not append. If append
# is preferable, modify this file accordingly.

###
# Generic settings for HADOOP
###

# Technically, the only required environment variable is JAVA_HOME.
# All others are optional. However, the defaults are probably not
# 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-11-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.
#
#

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos ^M-U Undo
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Linter ^_ Go To Line ^-E Redo
```

Keluar dan simpan dengan Ctrl + X → Y → Enter

#### 4. Edit berkas core-site.xml

Berkas core-site.xml berisi properties inti dari HDFS dan hadoop. Buka berkas core-site.xml dengan command:

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

Sampai bagian terakhir seperti pada gambar dibawah

```
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>  
</configuration>  
█
```

^G Get Help   ^O Write Out   ^W Where Is   ^K Cut Text   ^J Justify   ^C Cur Pos  
^X Exit   ^R Read File   ^\ Replace   ^U Uncut Text   ^T To Spell   ^\_ Go To Line

Tambahkan tulisan di bawah ini diantara <configuration> dan </configuration> untuk mengganti setting sistem terkait local file yang default

```
<property>  
  <name>hadoop.tmp.dir</name>  
  <value>/home/hduser/tmpdata</value>  
</property>  
  
<property>  
  <name>fs.default.name</name>  
  <value>hdfs://127.0.0.1:9000</value>  
</property>
```

Sehingga hasilnya seperti dibawah ini :

```
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
<property>  
  <name>hadoop.tmp.dir</name>  
  <value>/home/hduser/tmpdata</value>  
</property>  
<property>  
  <name>fs.default.name</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   ^C Cur Pos  
^X Exit   ^R Read File   ^\ Replace   ^U Uncut Text   ^T To Spell   ^\_ Go To Line

Keluar dan simpan dengan Ctrl + X → Y → Enter.



5. Keluar dari direktori saat ini dengan mengetikkan `cd`. Kemudian buat folder untuk data temporer pada lokasi `/home/hduser/tmpdata` seperti gambar di bawah ini.

```
hduser@bigdata-VirtualBox:~$ mkdir tmpdata
hduser@bigdata-VirtualBox:~$ ls
examples.desktop  hadoop-3.3.0  hadoop-3.3.0.tar.gz  text.txt  tmpdata
hduser@bigdata-VirtualBox:~$
```

6. Kemudian kembali ke direktori sebelumnya sesuai gambar di bawah ini, dan buat folder baru untuk direktori NameNode dan DataNode di dalam folder `hadoop`.

```
hduser@bigdata-VirtualBox:~$ cd hadoop-3.3.0/
hduser@bigdata-VirtualBox:~/hadoop-3.3.0$ mkdir dfsdata
hduser@bigdata-VirtualBox:~/hadoop-3.3.0$ ls
bin      etc      lib      LICENSE-binary  LICENSE.txt  NOTICE.txt  sbin
dfsdata  include  libexec  licenses-binary  NOTICE-binary  README.txt  share
hduser@bigdata-VirtualBox:~/hadoop-3.3.0$ cd dfsdata/
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/dfsdata$ mkdir datanode
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/dfsdata$ mkdir namenode
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/dfsdata$ ls
datanode  namenode
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/dfsdata$
```

7. Setting pada berkas `hdfs-site.xml`

Isi dari berkas `hdfs-site.xml` adalah untuk menyimpan metadata, berkas `fsimage` dan log dari edit file, serta untuk mendefinisikan direktori untuk NameNode dan DataNode. Kemudian, lakukan setting untuk jumlah replikasi, secara default maka ada 3 replikasi, karena hanya ada 1 node, maka ubah nilai replikasi ini di berkas `hdfs-site.xml` juga. Buka berkas `hdfs-site.xml` dengan command berikut.

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

```
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>
</configuration>
█

^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File  ^_ Replace   ^U Uncut Text ^T To Spell  ^_ Go To Line
```

Tambahkan perintah dibawah ini diantara `<configuration>` dan `</configuration>`, isi path folder sesuai dengan lokasi direktori yang sudah dibuat sebelumnya. Untuk `dfs.replication` akan diisi dengan 1.

```

<property>

  <name>dfs.namenode.data.dir</name>

  <value>/home/hduser/hadoop-3.3.0/dfsdata/namenode</value>

</property>

<property>

  <name>dfs.datanode.data.dir</name>

  <value>/home/hduser/hadoop-3.3.0/dfsdata/datanode</value>

</property><property>

  <name>dfs.replication</name>

  <value>1</value>

</property>

```

```

hduser@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 /home/hduser/hadoop-3.3.0/etc/hadoop/hdfs-site.xml Modified
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->

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

^G Get Help  ^O Write Out  ^W Where Is   ^K Cut Text   ^J Justify    ^C Cur Pos   M-U Undo
^X Exit      ^R Read File  ^\ Replace    ^U Uncut Text ^T To Spell   ^_ Go To Line M-E Redo

```

Keluar dan simpan dengan Ctrl + X → Y → Enter.

## 8. Pengaturan berkas mapred-site.xml

Buka dengan command

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

```
?xml version="1.0"?>
<?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 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>
</configuration>
```

Read 21 lines

Get Help Write Out Where Is Cut Text Justify Cur Pos  
Exit Read File Replace Uncut Text To Spell Go To Line

Tambahkan perintah dibawah ini untuk mengganti konfigurasi default dari MapReduce Framework ke YARN.

```
<property>

  <name>mapreduce.framework.name</name>

  <value>yarn</value>

</property>
```

```
hduser@bigdata-VirtualBox: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 /home/hduser/hadoop-3.2.1/etc/hadoop/mapred-site.xml Modified

<?xml version="1.0"?>
<?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 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>
```

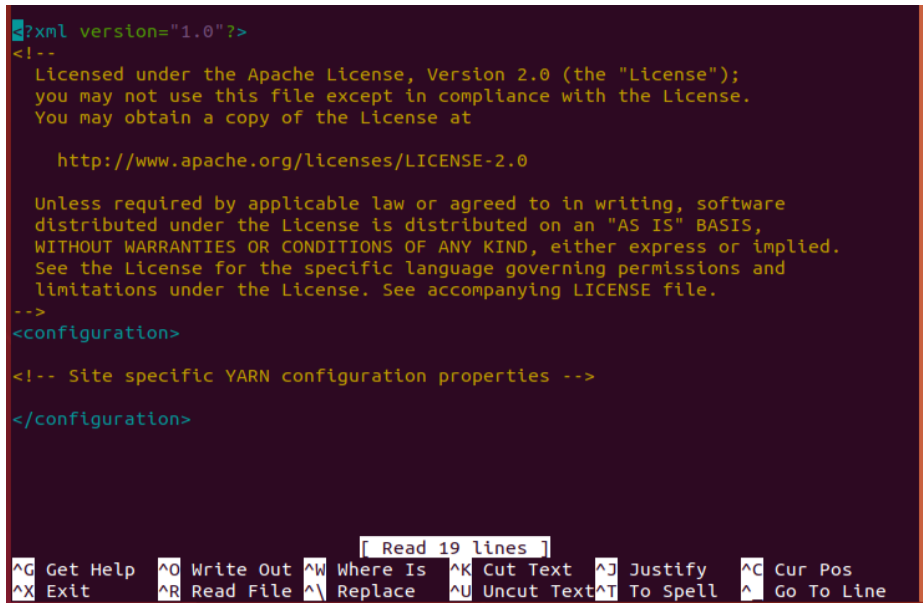
Get Help Write Out Where Is Cut Text Justify Cur Pos  
Exit Read File Replace Uncut Text To Spell Go To Line

Keluar dan simpan dengan Ctrl + X → Y → Enter.

## 9. Pengaturan berkas yarn-site.xml

Berkas yarn-site.xml adalah tempat untuk mendefinisikan pengaturan terkait YARN. Berisi konfigurasi untuk nodemanager, resourcemanager, container dan application master. Buka berkas yarn-site.xml dengan perintah

```
sudo nano $HADOOP_HOME/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>

  <!-- Site specific YARN configuration properties -->

</configuration>
```

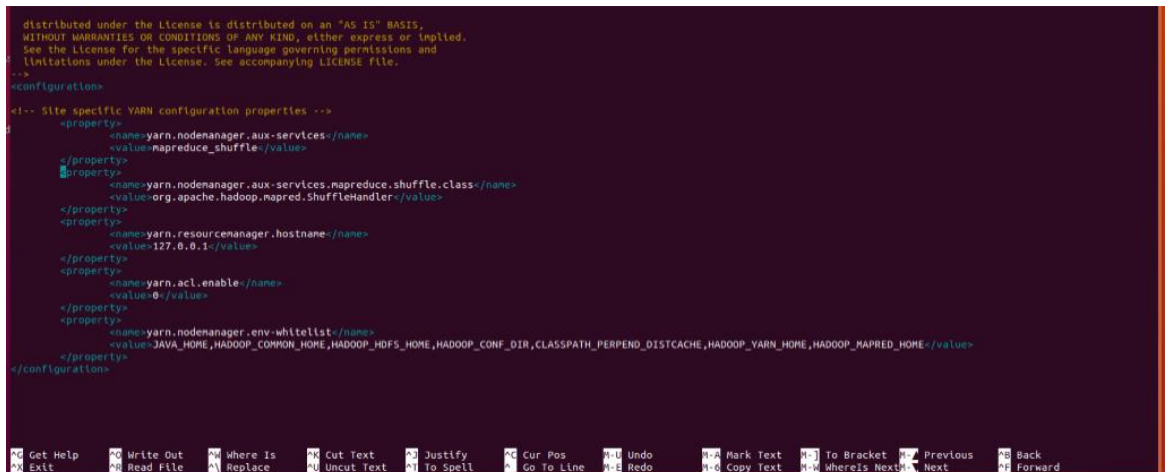
Tambahkan konfigurasi sebagai berikut diantara `<configuration>` dan `</configuration>`:

```
<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,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
```

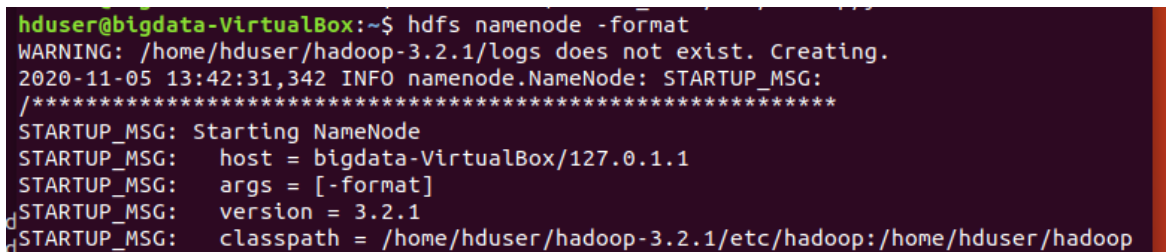
</property>



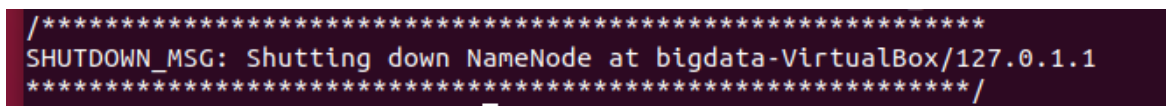
Keluar dan simpan dengan Ctrl + X → Y → Enter.

## 10. Format HDFS nameNode dengan perintah

hdfs namenode -format



Jika sudah selesai, maka ada pesan shutdown



## G. Memulai Cluster Hadoop

1. Pindah ke folder hadoop-3.3.0/sbin

```
hduser@bigdata-VirtualBox:~$ cd hadoop-3.3.0/sbin
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$
```

2. Jalan perintah untuk memulai NameNode dan DataNode  
./start-dfs.sh

Tunggu prosesnya hingga selesai.

```
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bigdata-VirtualBox]
```

Jika ada warning, bisa diabaikan

```
bigdata-VirtualBox: Warning: Permanently added 'bigdata-virtualbox' (ECDSA) to the
list of known hosts.
```

3. Menjalankan YARN dengan command berikut.  
./start-yarn.sh

```
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$
```

4. Cek daemon apakah sudah berjalan dengan command berikut.  
jps

```
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$ jps
11040 NodeManager
10322 DataNode
11122 Jps
10870 ResourceManager
10553 SecondaryNameNode
10156 NameNode
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$
```

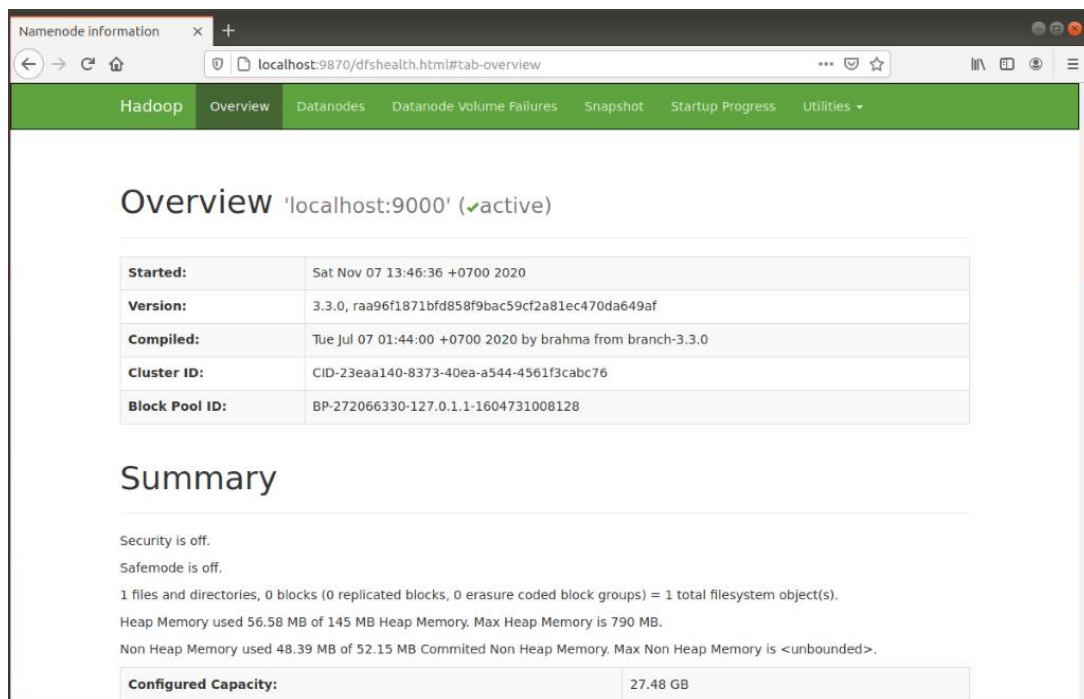
Apabila NameNode tidak terdefiniskan pada command `jps` seperti gambar di bawah, langkah yang harus dilakukan adalah menjalankan command `./stop-all.sh`

```
hduser@nurlaita-VirtualBox:~/hadoop-3.3.0/sbin$ jps
11776 SecondaryNameNode
12209 NodeManager
12580 Jps
11546 DataNode
12044 ResourceManager
```

Tunggu hingga proses selesai, kemudian kembali ke langkah nomor F-10 dengan mengetikkan kembali command `hadoop namenode -format`. Kemudian lanjutkan ke tahapan G-2 untuk memulai Cluster Hadoop hingga selesai.

## H. Akses Hadoop UI dari Web Browser

1. Buka NameNode pada *browser* di Ubuntu dengan mengetikkan alamat berikut.  
`localhost:9870`





2. Buka DataNode pada *browser* di Ubuntu dengan mengetikkan alamat berikut.  
localhost:9864/datanode.html

**DataNode on bigdata-VirtualBox:9866**

<b>Cluster ID:</b>	CID-23eaa140-8373-40ea-a544-4561f3cab76
<b>Version:</b>	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af

**Block Pools**

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-272066330-127.0.1.1-1604731008128	RUNNING	1s	3 minutes	0 B (128 MB)

**Volume Information**

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/hduser/hadoop-	DISK	28 KB	11.85 GB	0 B	0 B	0

3. Buka YARN pada *browser* di Ubuntu dengan mengetikkan alamat berikut.  
localhost:8088/cluster

**hadoop**

**Cluster**

- About
- Nodes
- Node Labels
- Applications
  - NEW
  - NEW SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler

**Cluster Metrics**

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

**Cluster Nodes Metrics**

Active Nodes	Decommissioning Nodes	Decommissioned
1	0	0

**Scheduler Metrics**

Scheduler Type	Scheduling Resource Type
Capacity Scheduler	[memory-mb (unit=Mb), vcores]

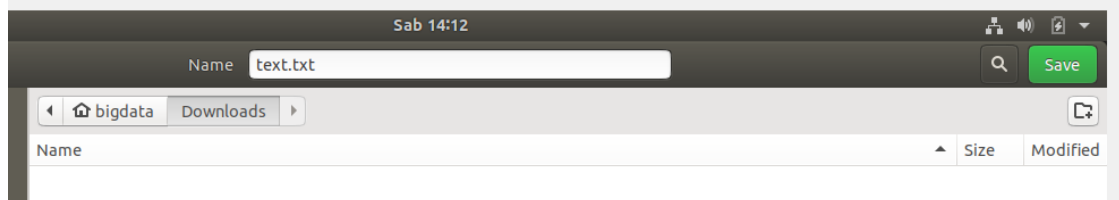
Showing 0 to 0 of 0 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	Fin
----	------	------	------------------	------------------	-------	----------------------	-----------	------------	-----

Showing 0 to 0 of 0 entries

## 1. Memindahkan Data dari dan ke HDFS

1. Download file txt berikut pada Ubuntu yang bersumber dari link berikut:  
<https://ocw.mit.edu/ans7870/6/6.006/s08/lecturenotes/files/t8.shakespeare.txt>
2. Simpan di folder download dengan cara klik kanan → save as, dan beri nama text.txt



3. Mengecek apakah file sudah tersedia pada folder download dengan cara seperti gambar di bawah ini. Gunakan user asal masing-masing (bukan hduser).

```
bigdata@bigdata-VirtualBox: ~/Downloads
File Edit View Search Terminal Help
(base) bigdata@bigdata-VirtualBox:~$ ls
anaconda3  Documents  examples.desktop  Pictures  Templates
Desktop    Downloads  Music             Public    Videos
(base) bigdata@bigdata-VirtualBox:~$ cd Downloads/
(base) bigdata@bigdata-VirtualBox:~/Downloads$ ls
Anaconda3-2020.07-Linux-x86_64.sh  spark-3.0.1-bin-hadoop2.7.tgz  text.txt
(base) bigdata@bigdata-VirtualBox:~/Downloads$
```

4. Pindah berkas text.txt ke home hduser, terminal dengan user big data menggunakan command berikut.

```
sudo cp /home/bigdata/Downloads/text.txt /home/hduser/text.txt
```

```
(base) bigdata@bigdata-VirtualBox:~$ sudo cp /home/bigdata/Downloads/text.txt /home/hduser/text.txt
```

5. Dengan terminal user hduser, cek apakah sudah ada berkasnya menggunakan command berikut.

```
hduser@bigdata-VirtualBox:~$ ls -l
total 494372
-rw-r--r--  1 hduser  hadoop      8980 Nov  4 07:56 examples.desktop
drwxr-xr-x 12 hduser  hadoop      4096 Nov  7 13:36 hadoop-3.3.0
-rwxr-x---  1 hduser  root    500749234 Nov  7 13:16 hadoop-3.3.0.tar.gz
-rw-r--r--  1 root    root    5458199 Nov  7 14:14 text.txt
drwxr-xr-x  4 hduser  hadoop      4096 Nov  7 13:37 tmpdata
hduser@bigdata-VirtualBox:~$
```

6. Change owner dari berkas ke hduser menggunakan command berikut.

```
Sudo chown hduser text.txt
```

```
hduser@bigdata-VirtualBox:~$ sudo chown hduser text.txt
[sudo] password for hduser:
hduser@bigdata-VirtualBox:~$ ls -l
total 494372
-rw-r--r--  1 hduser  hadoop      8980 Nov  4 07:56 examples.desktop
drwxr-xr-x 12 hduser  hadoop     4096 Nov  7 13:36 hadoop-3.3.0
-rwxr-x---  1 hduser  root    500749234 Nov  7 13:16 hadoop-3.3.0.tar.gz
-rw-r--r--  1 hduser  root    5458199 Nov  7 14:14 text.txt
drwxr-xr-x  4 hduser  hadoop     4096 Nov  7 13:37 tmpdata
hduser@bigdata-VirtualBox:~$
```

7. Buat folder pada HDFS menggunakan command berikut.

```
hadoop fs -mkdir /user
```

```
hadoop fs -mkdir /user/modul8
```

```
hduser@bigdata-VirtualBox:~$ hadoop fs -mkdir /user
hduser@bigdata-VirtualBox:~$ hadoop fs -mkdir /user/modul6
```

8. Copy berkas ke HDFS menggunakan command berikut.

```
hadoop fs -copyFromLocal text.txt /user/modul8
```

```
hduser@bigdata-VirtualBox:~$ hadoop fs -copyFromLocal /home/hduser/text.txt /user/modul6
```

9. Cek Apakah berhasil di-copy atau tidak. Jika berhasil akan menemukan 1 item tersebut.

```
hadoop fs -ls /user/modul8
```

```
hduser@bigdata-VirtualBox:~$ hadoop fs -ls /user/modul6
Found 1 items
-rw-r--r--  1 hduser  supergroup    5458199 2020-11-07 15:23 /user/modul6/text.txt
```

10. Copy data dari HDFS ke lokal menggunakan command berikut.

```
hadoop fs -copyToLocal /user/modul8/text.txt
```

```
/home/hduser/textkopi.txt
```

```
hduser@bigdata-VirtualBox:~$ hadoop fs -copyToLocal /user/modul6/text.txt /home/hduser/textkopi.txt
hduser@bigdata-VirtualBox:~$ ls
examples.desktop  hadoop-3.3.0  hadoop-3.3.0.tar.gz  textkopi.txt  text.txt  tmpdata
hduser@bigdata-VirtualBox:~$
```