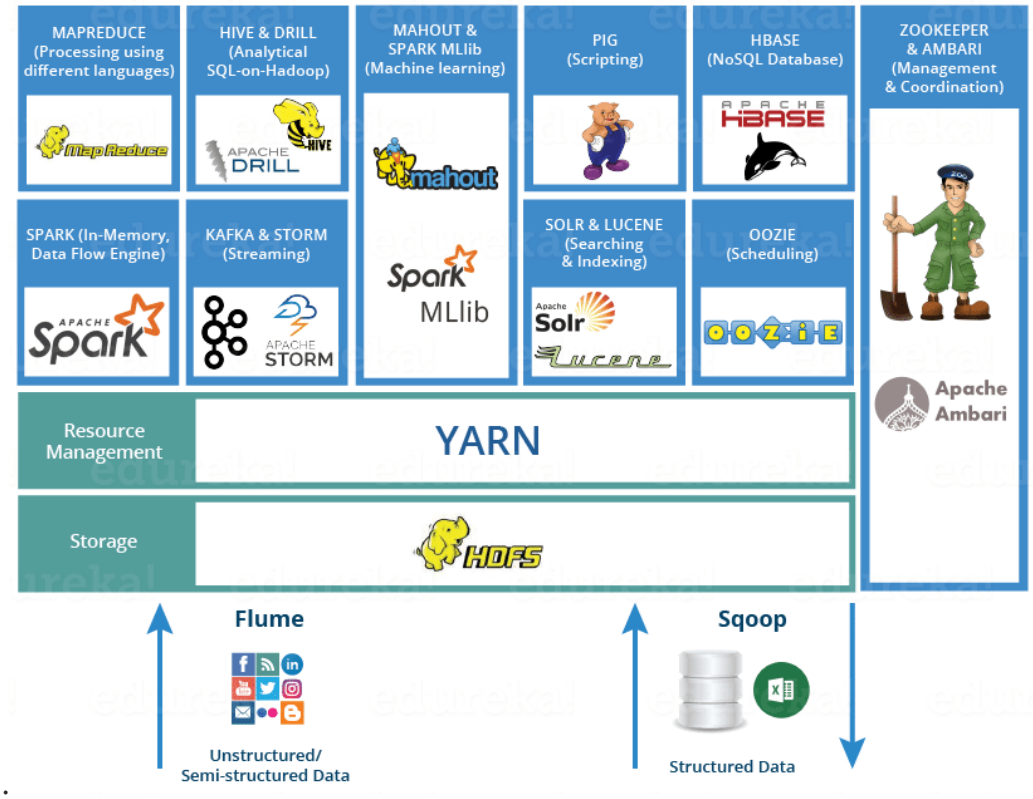
Modul 7

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.

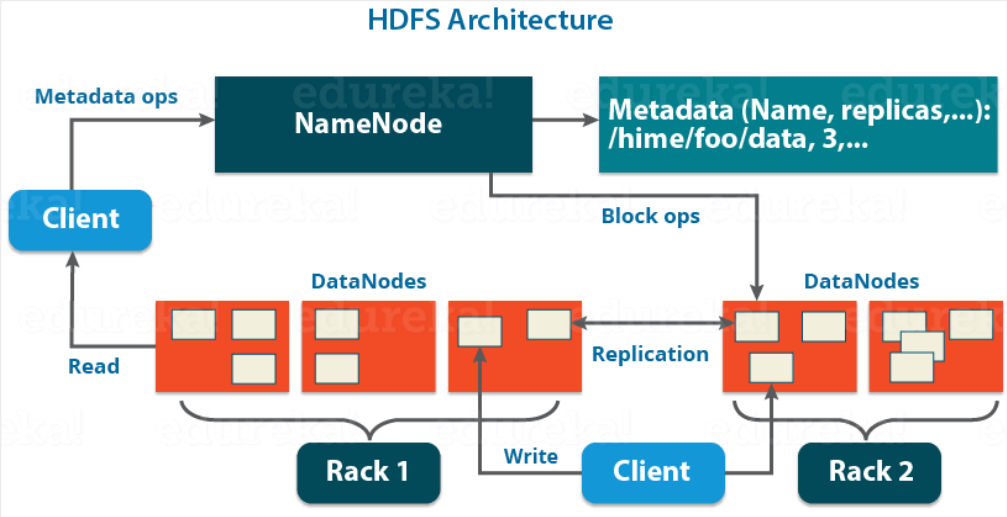


Gambar Ekosistem Hadoop Pada Umumnya

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

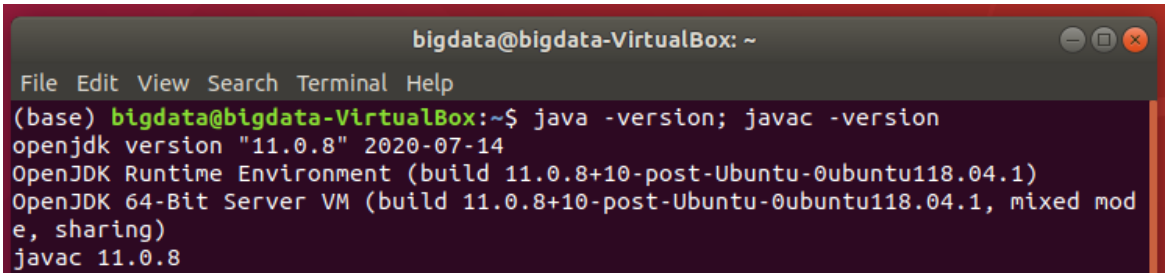
A. Mengecek Versi Java

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

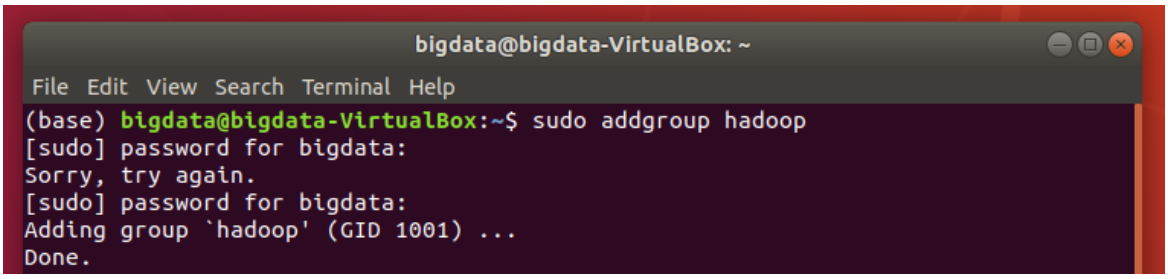
   → `java -version`

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



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`



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.

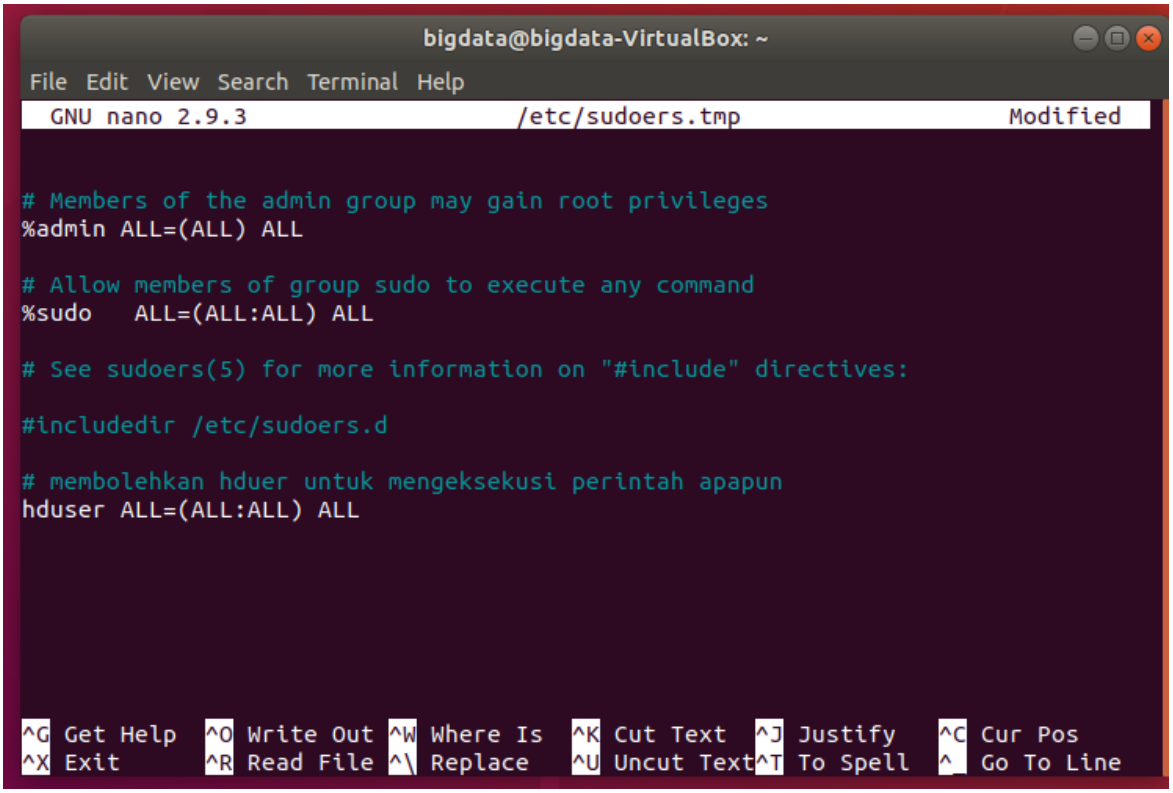
```
(base) bigdata@bigdata-VirtualBox:~$ sudo adduser -ingroup hadoop hduser
Adding user `hduser' ...
Adding new user `hduser' (1001) with group `hadoop' ...
Creating home directory `/home/hduser' ...
Copying files from `/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password 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

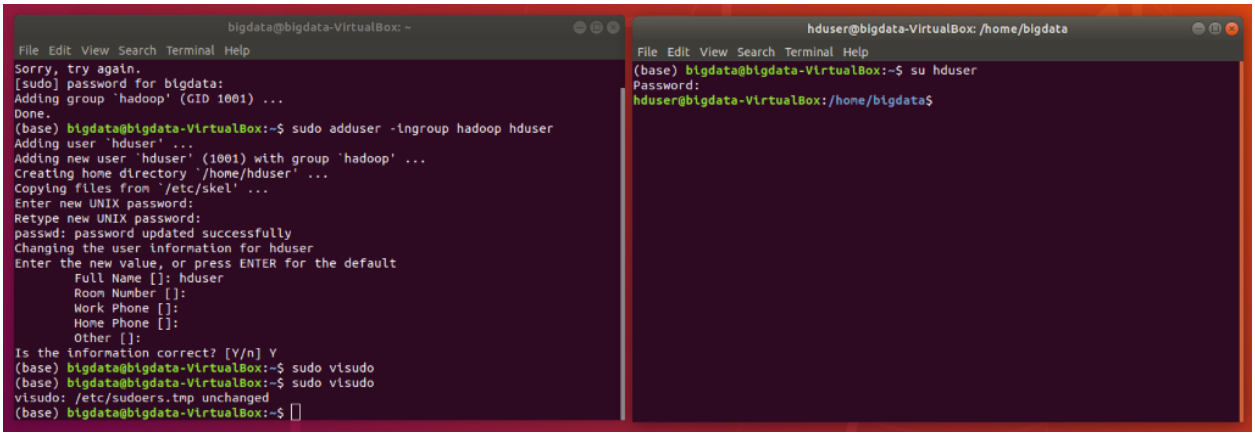
   hduser ALL=(ALL:ALL) ALL

   Sesuaikan dengan gambar di bawah ini.



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

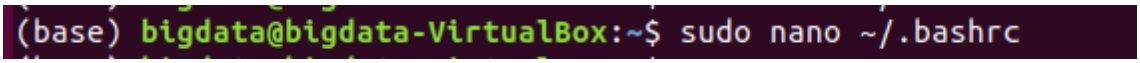
4. Buka terminal baru dan masuk sebagai user hduser (lihat terminal yang kanan). Masukkan password sesuai yang telah diinput pada langkah nomor dua.
   `su hduser`



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

C. Setting Java Environment

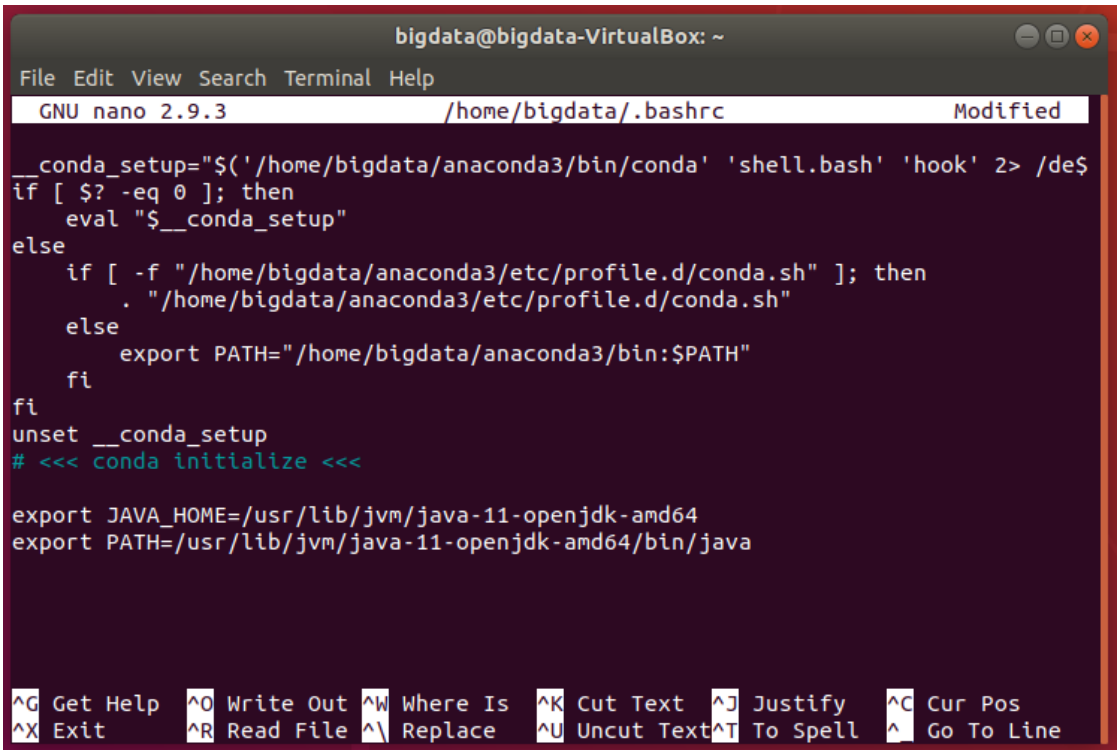
1. Kembali gunakan **terminal root** dan ubah .bash dengan command berikut.
   `sudo nano ~/.bashrc`



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



   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;
```

…

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

```
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-cli
ent -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-c
lient -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-c
lient
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-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
```

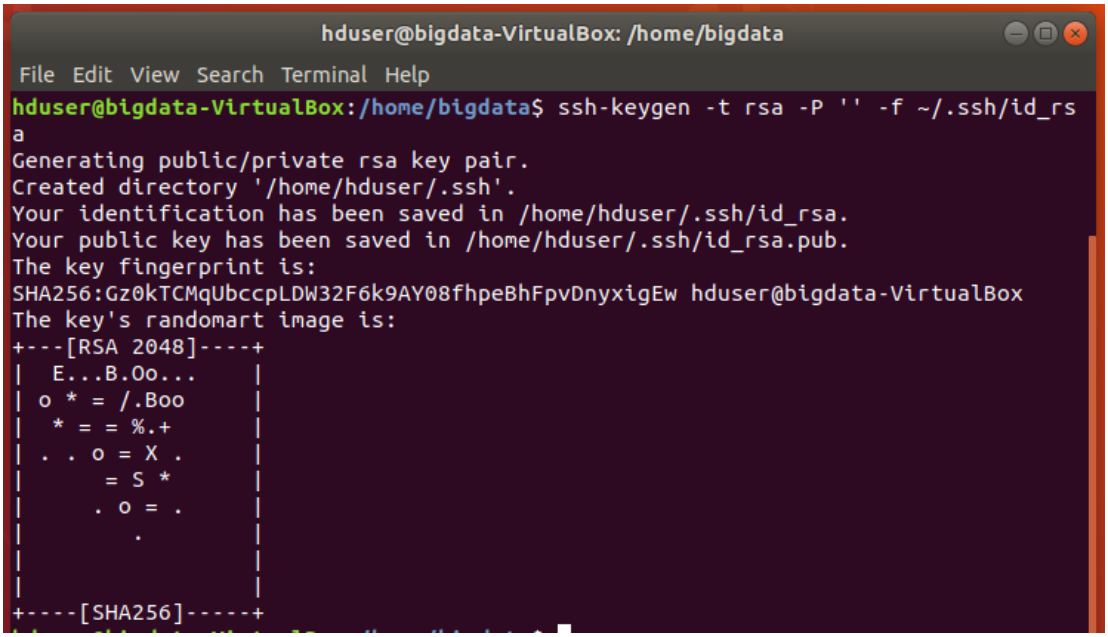
```
fawwaz@fawwaz-VirtualBox:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:7.6p1-4ubuntu0.3).
openssh-server is already the newest version (1:7.6p1-4ubuntu0.3).
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
```

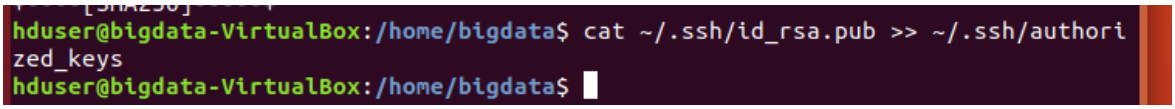
3. Pindah ke user hduser, atau ke terminal yang sebelumnya sudah dibuka dengan user hduser. Kemudian Generate SSH key pair dan definisikan lokasinya untuk dapat disimpan menggunakan command berikut.

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



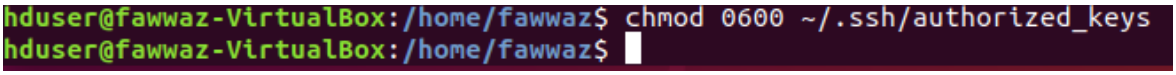
4. Gunakan command berikut untuk menyimpan public key sebagai authorized_keys dalam direktori ssh.

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



5. Set hak akses untuk user menggunakan command berikut.

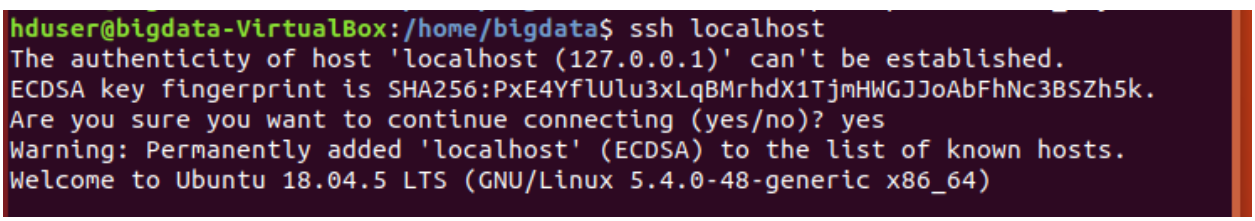
```
chmod 0600 ~/.ssh/authorized_keys
```



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.



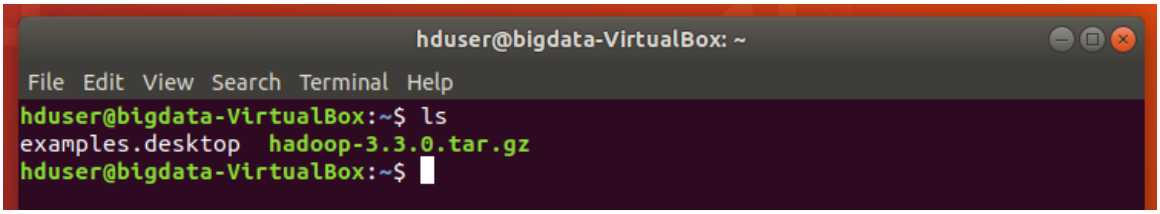
(opsional) !!! Jika masih meminta password, dapat mengetik command berikut.

```
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`

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

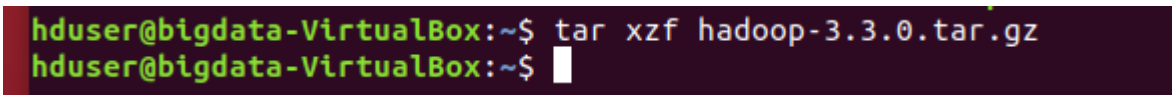


(opsional) !!! Jika mendownload secara terpisah, maka pastikan owner dari berkas adalah hduser dengan perintah

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

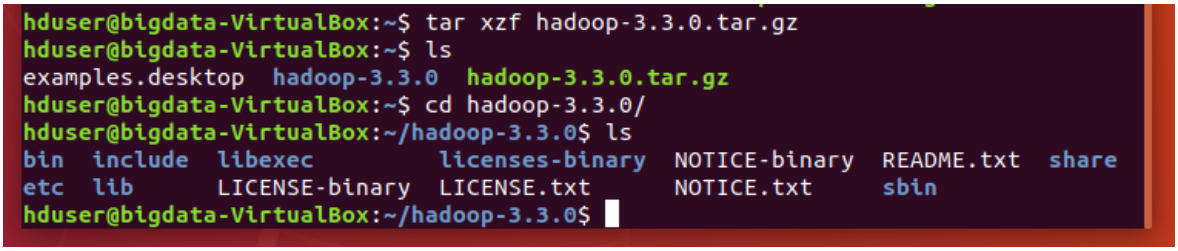
2. Extrak hasil download menggunaan command berikut.
   `tar xzf hadoop-3.3.0.tar.gz`

Tunggu hingga proses ekstraksi selesai.



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

   Kemudian command ls.

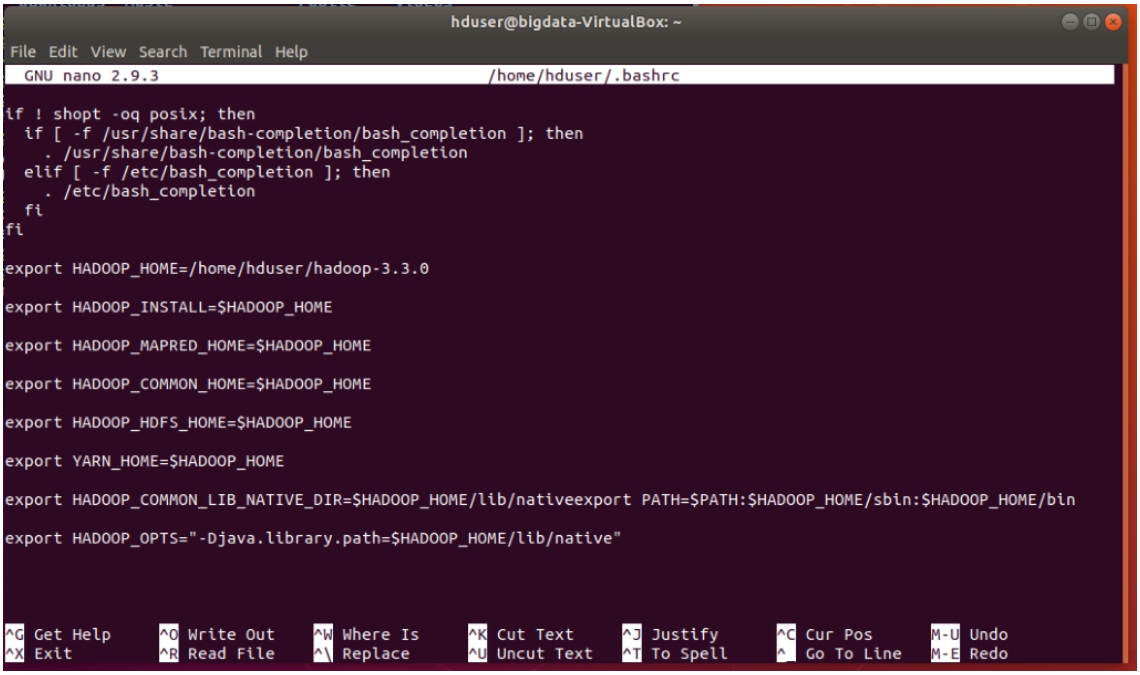


4. **DISARANKAN melakukan SNAPSHOT** ketika sudah sampai pada langkah ini untuk mempermudah mengatasi *error* yang kemungkinan akan terjadi kedepannya.

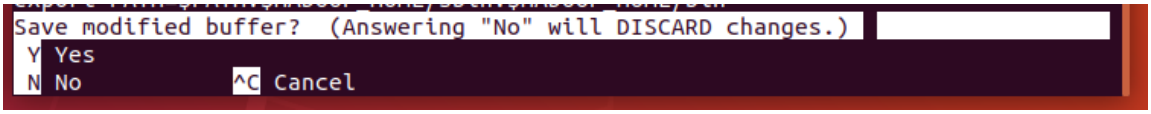
F. Setting Hadoop dengan Single Node / Psedo Distributed Mode

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:

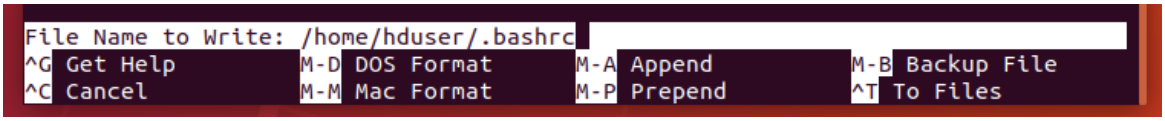
```
export HADOOP_HOME=/home/hduser/hadoop-3.3.0

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=$HADOOP_HOME/lib/nativeexport
PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin

export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```



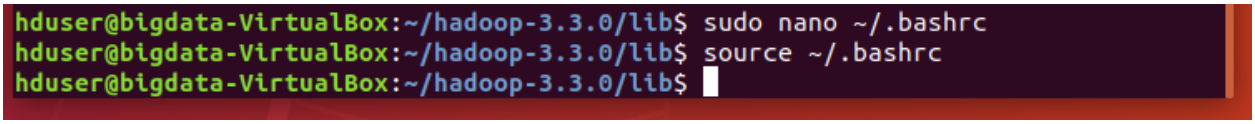
Kemudian control x untuk keluar, Save dengan tuliskan Y



untuk file name biarkan sama. Langsung enter saja.

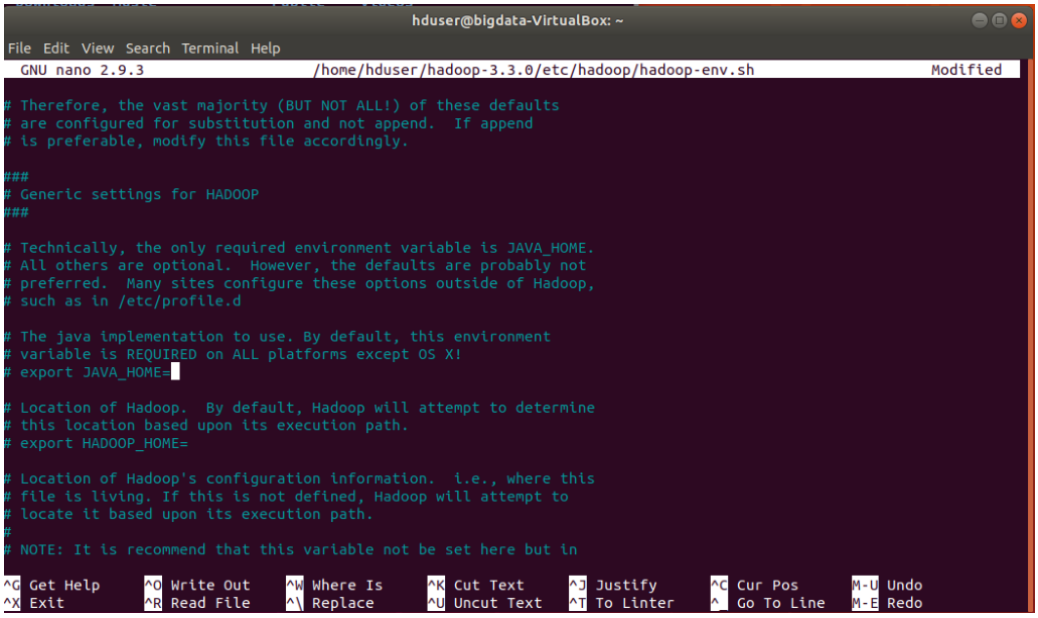


2. Jalankan perintah untuk menyimpan setting untuk environtmen yang berjalan dengan menggunakan perintah :
`source ~/.bashrc`



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

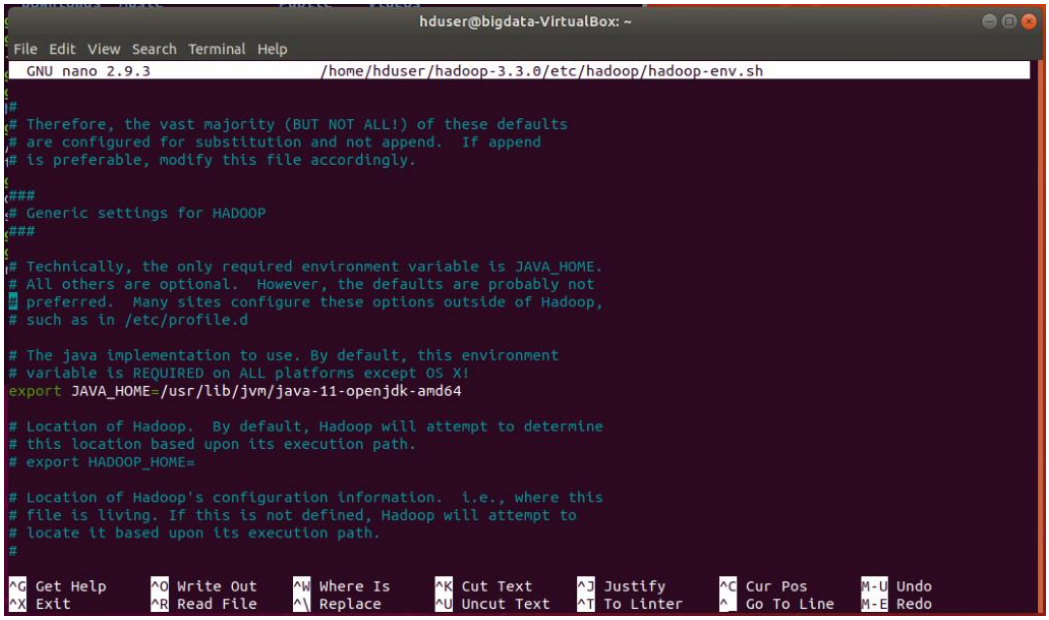
3. Pengeditan berkas hadoop-env.sh dengan command:
`sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh`



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
```

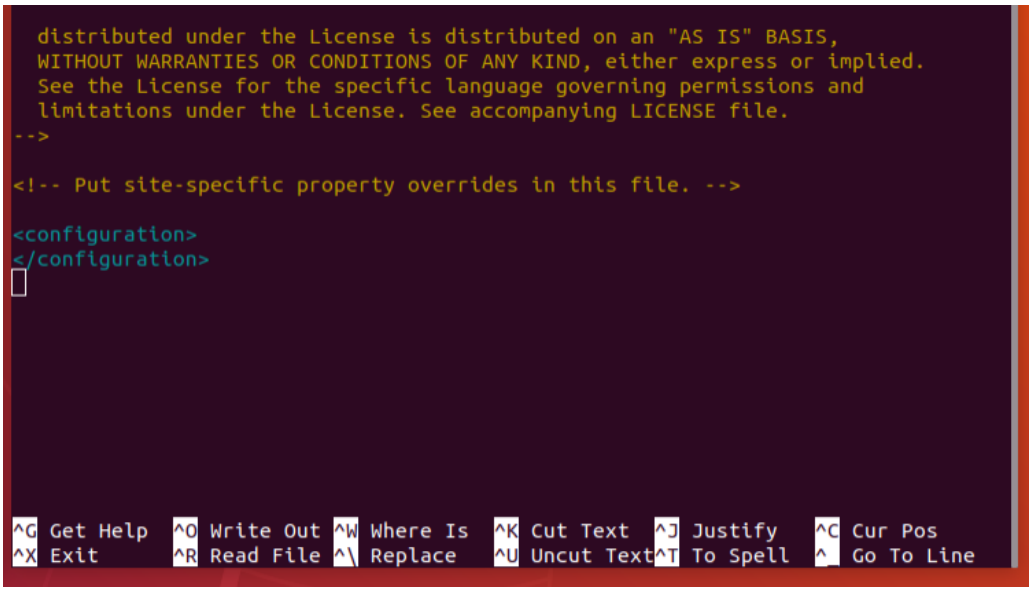


Keluar dan simpan dengan Ctrl + X → Y → Enter

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
   ```

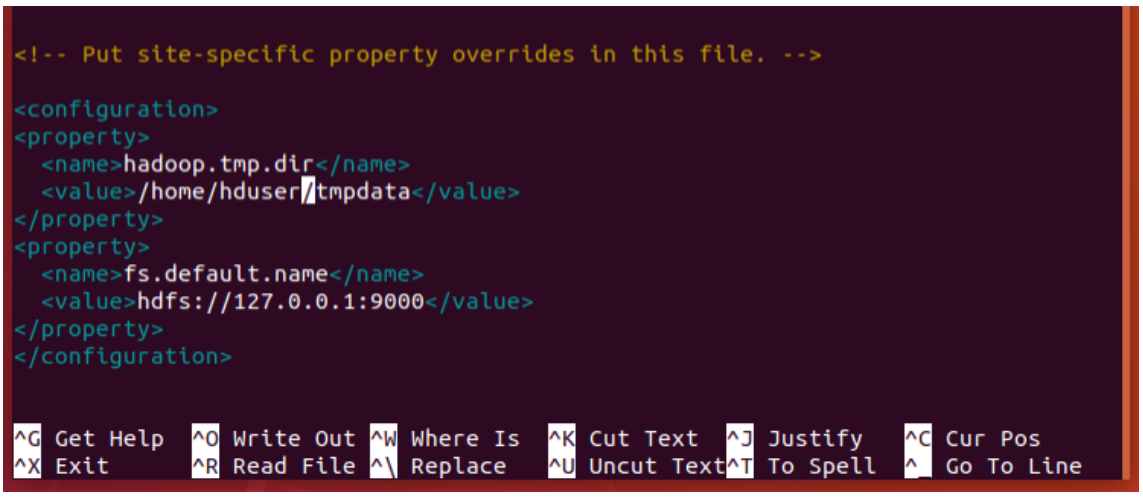
Sampai bagian terakhir seperti pada gambar dibawah



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

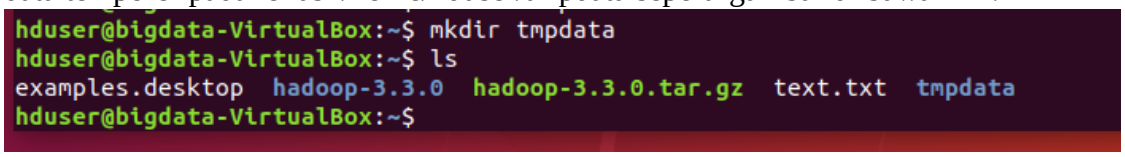
```
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/hduser/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
```

Sehingga hasilnya seperti dibawah ini :

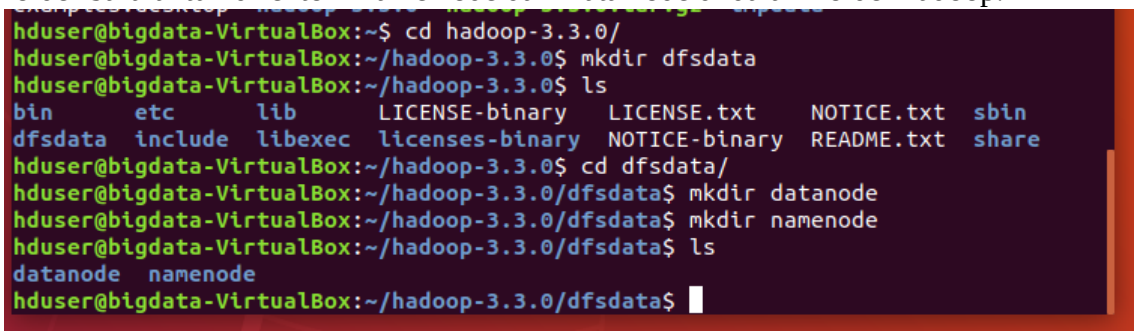


Keluar dan simpan dengan Ctrl + X → Y → Enter.

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.



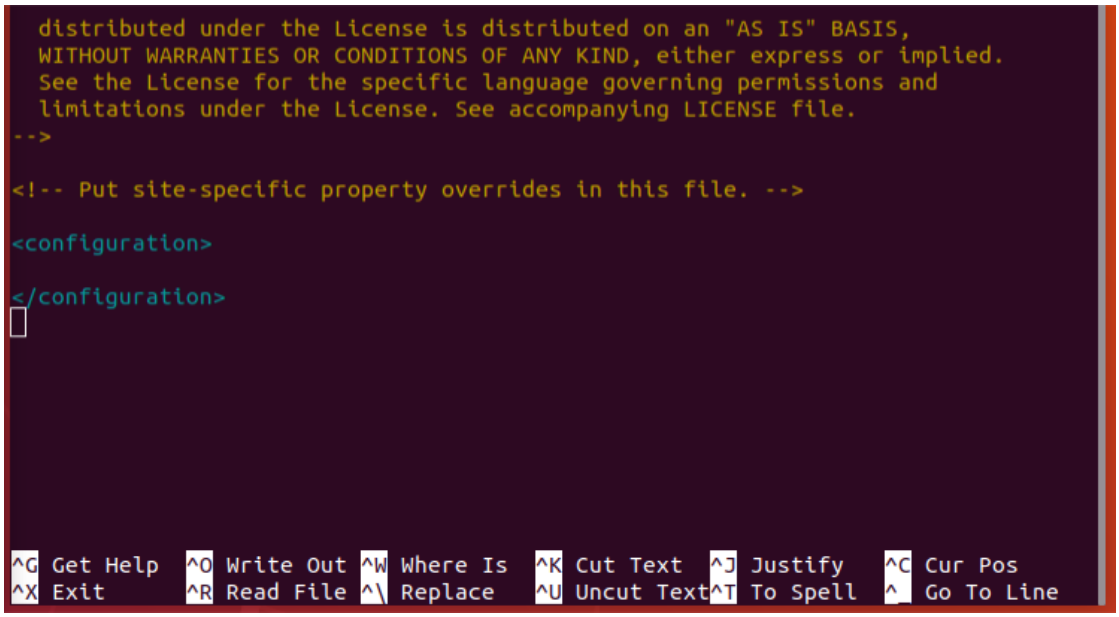
6. Kemudian kembali ke direktori sebelumnya sesuai gambar di bawah ini, dan buat folder baru untuk direktori NameNode dan DataNode di dalam folder hadoop.



7. Setting pada berkas hdfs-site.xml

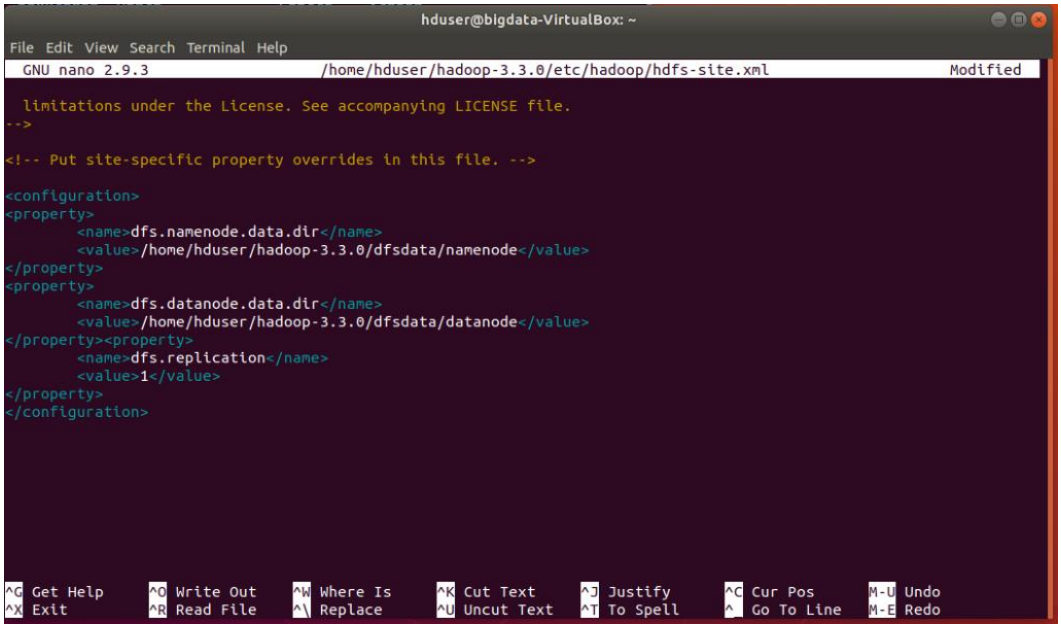
   Isi dari berkas hdfs-site.xml adalah untuk menyimpan metadata, berkas fsimage dan log dari edit file, serta untuk mendefinisikan direktori untuk NameNode dan DataNode. Kemudian, lakukan setting untuk jumlah replikasi, secara default maka ada 3 replikasi, karena hanya ada 1 node, maka ubah nilai replikasi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

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



   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>
```



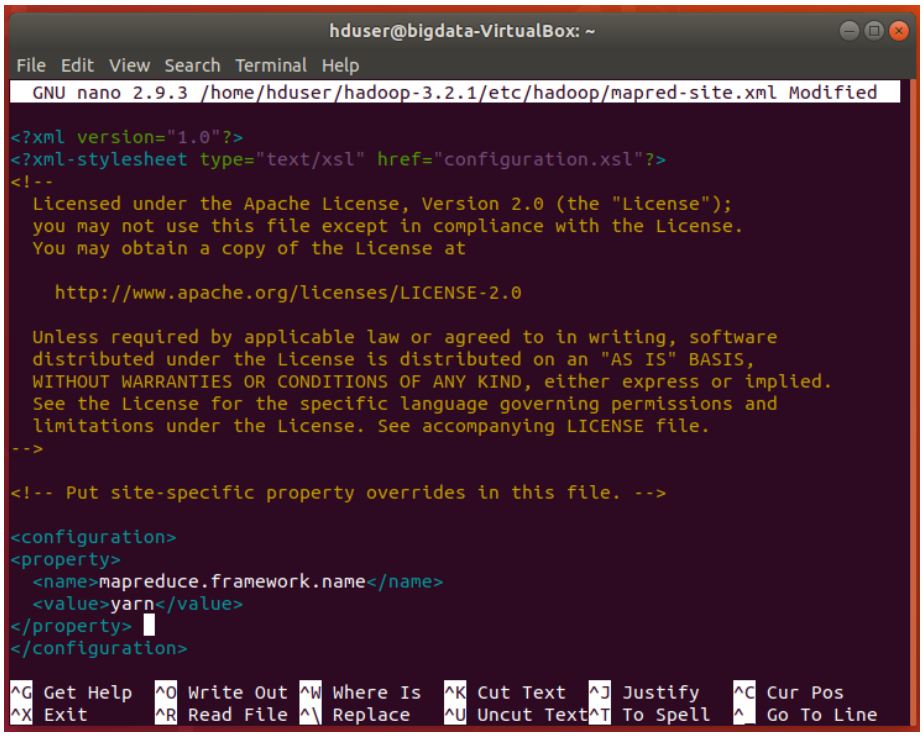
Keluar dan simpan dengan Ctrl + X → Y → Enter.

8. Pengaturan berkas mapred-site.xml
   Buka dengan command

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

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

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

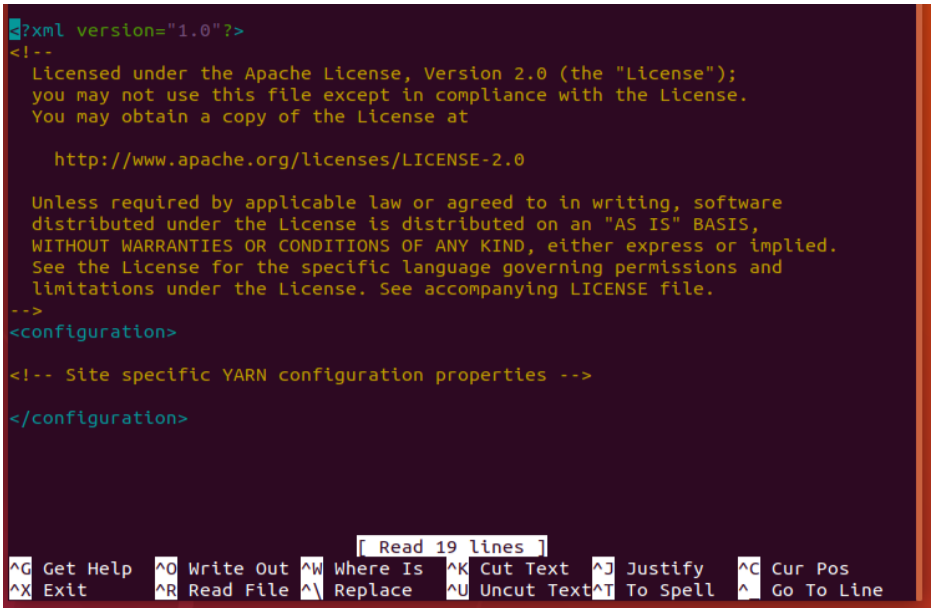


Keluar dan simpan dengan Ctrl + X → Y → Enter.

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

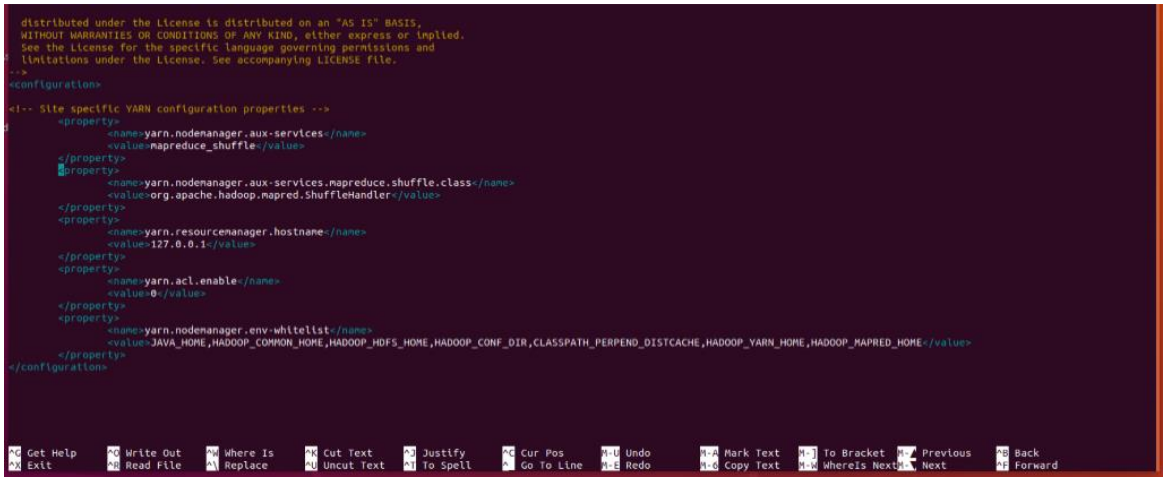
```
sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```



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

```
<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>
```

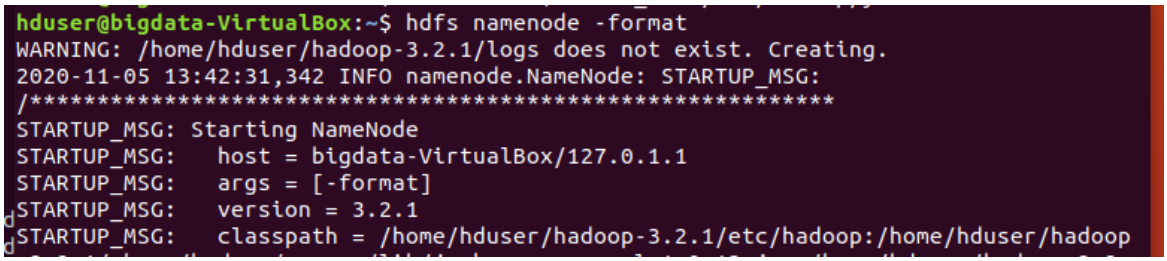
```
<property>

    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CO
    NF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAP
    RED_HOME</value>

</property>
```



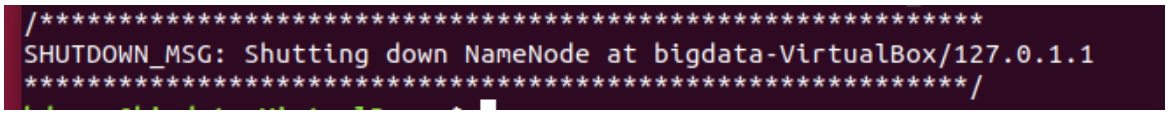
Keluar dan simpan dengan Ctrl + X → Y → Enter.

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
```

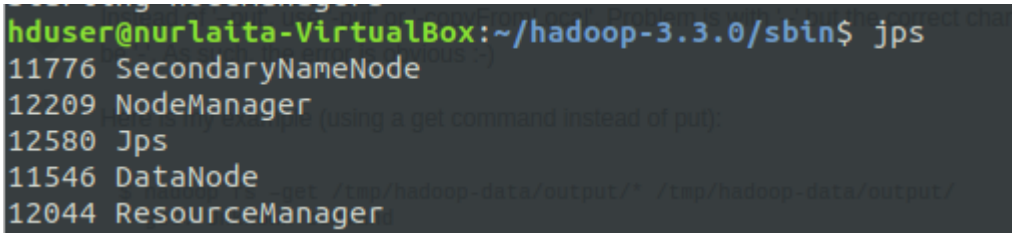
```
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hduser@bigdata-VirtualBox:~/hadoop-3.3.0/sbin$
```

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$
```

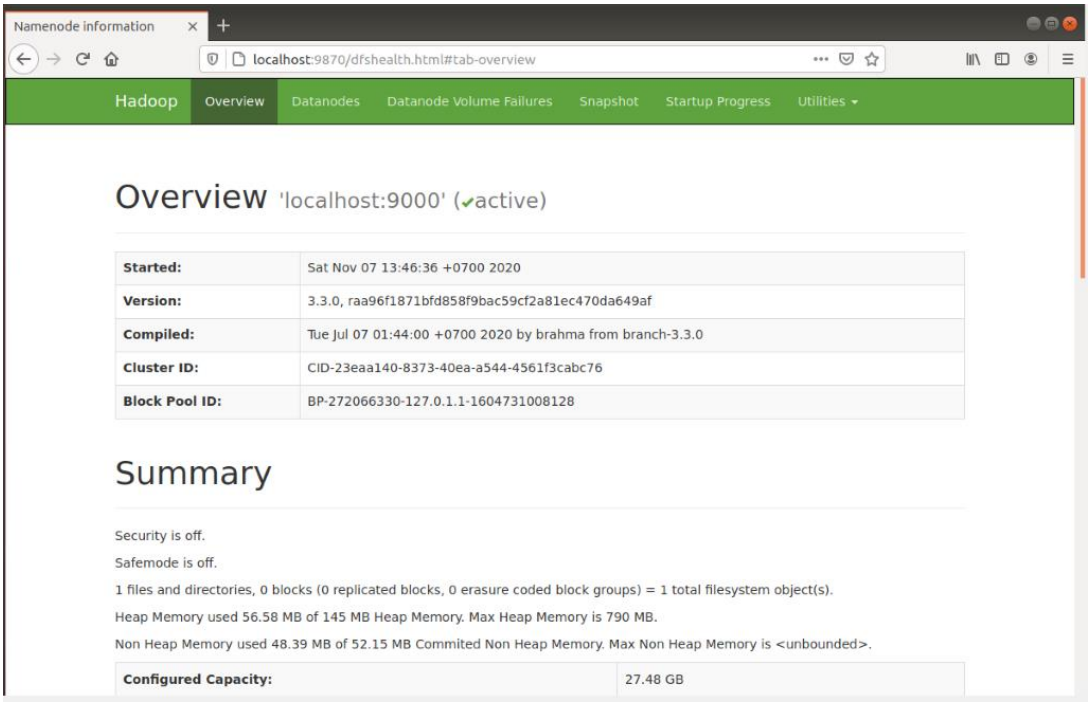
Apabila NameNode tidak terdefinisikan pada command jps seperti gambar di bawah, langkah yang harus dilakukan adalah menjalankan command `./stop-all.sh`



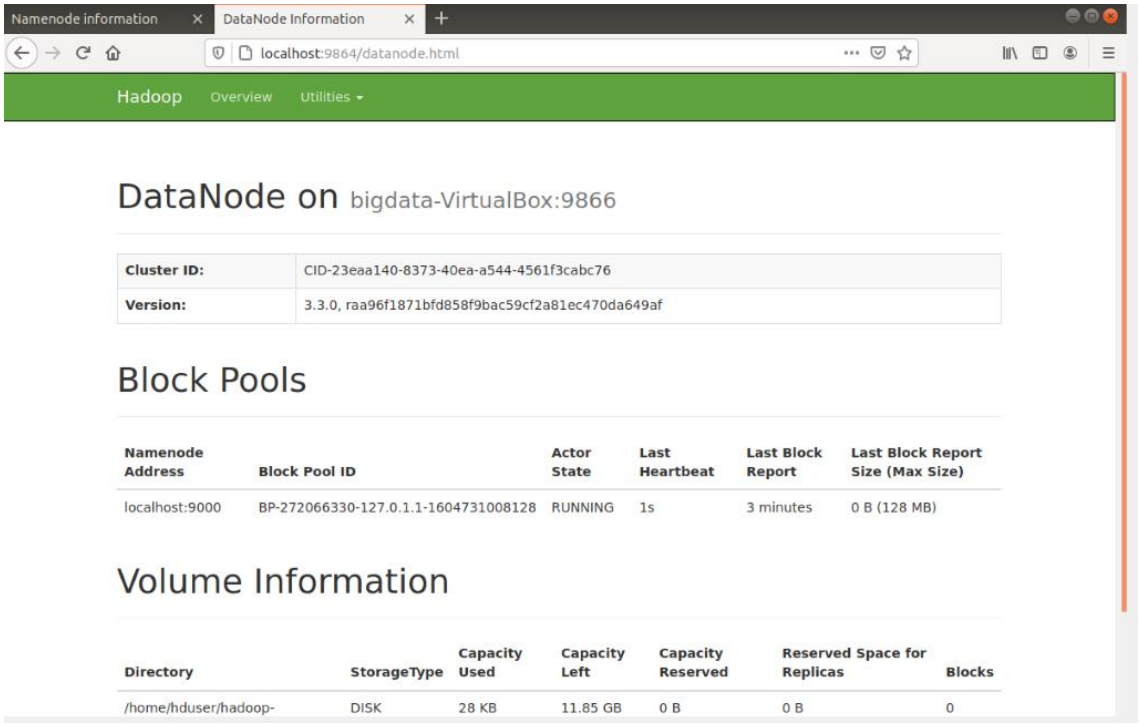
Tunggu hingga proses selesai, kemudian kembali ke langkah nomor F-10 dengan mengetikkan kembali command `hadoop namenode -format`. Kemudian lanjutkan ke tahapan G-2 untuk memulai Cluster Hadoop hingga selesai.

H. Akses Hadoop UI dari Web Browser

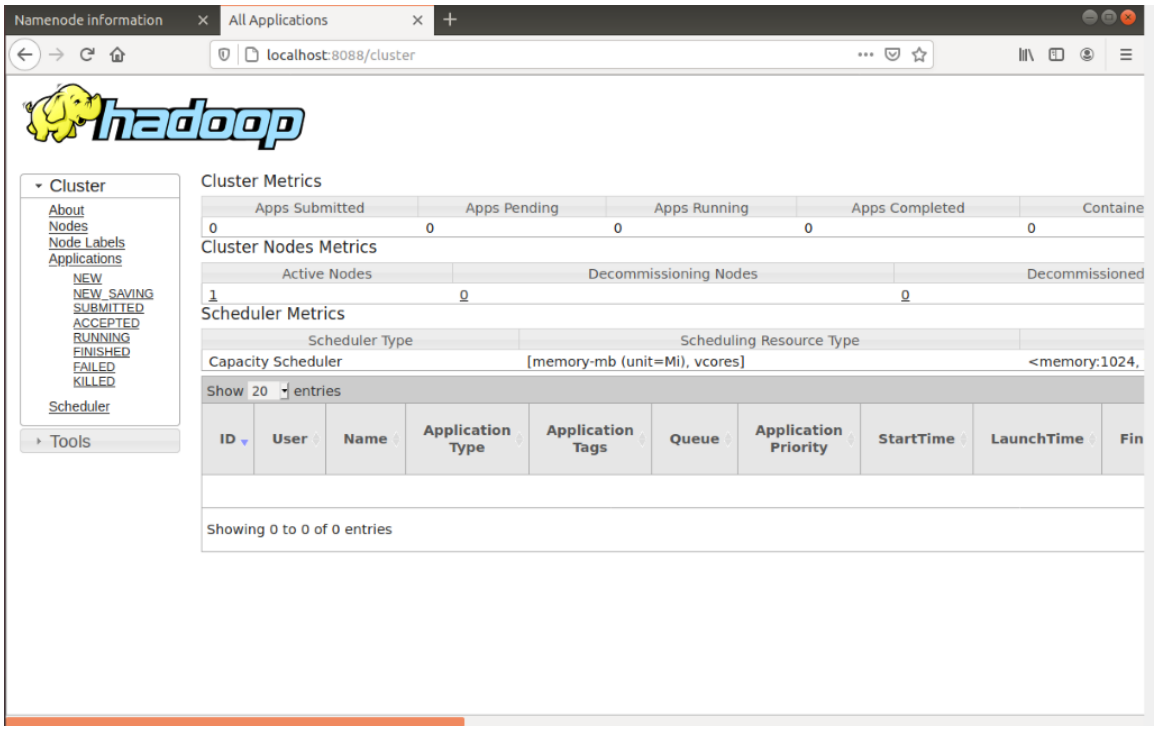
1. Buka NameNode pada *browser* di Ubuntu dengan mengetikkan alamat berikut. localhost:9870



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



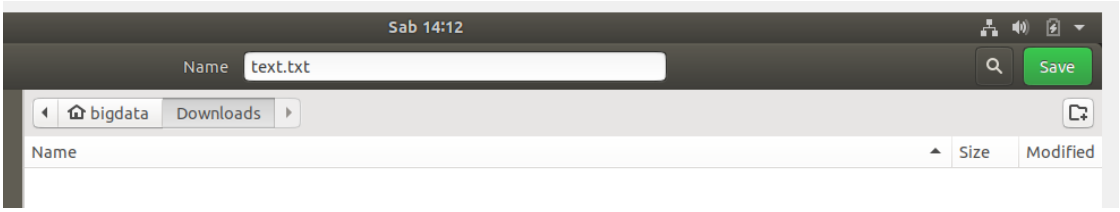
3. Buka YARN pada *browser* di Ubuntu dengan mengetikkan alamat berikut. localhost:8088/cluster



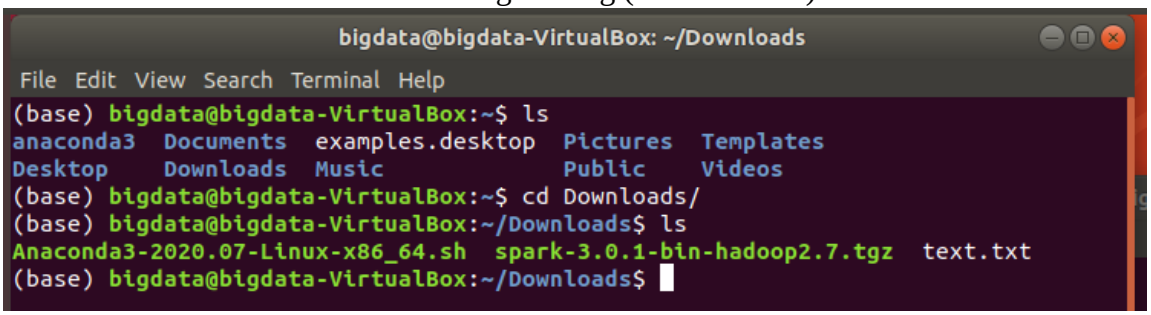
I. Memindahkan Data dari dan ke HDFS

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

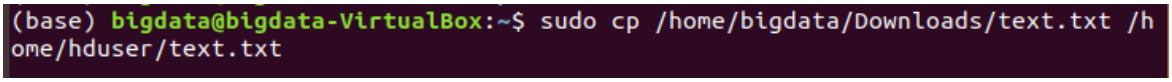


3. Mengecek apakah file sudah tersedia pada folder download dengan cara seperti gambar di bawah ini. Gunakan user asal masing-masing (bukan hduser).

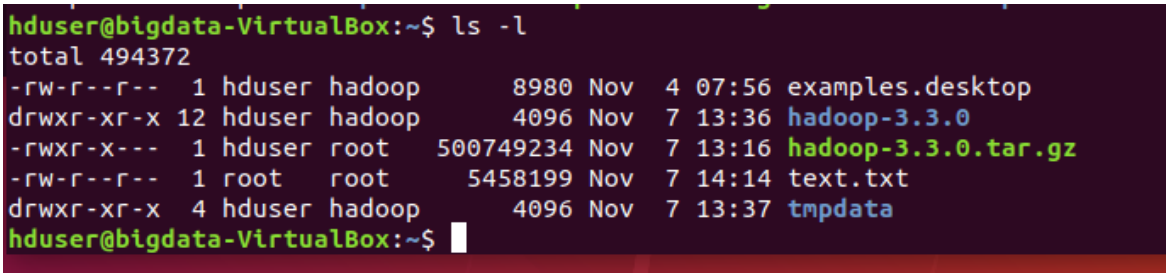


4. Pindah berkas text.txt ke home hduser, terminal dengan user big data menggunakan command berikut.

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



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



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:~$
```