- Install VirtualBox
- Install Ubuntu 20

Download di dalam Ubuntu::

1. com.oracle.JRE64-1.8.0.91-AutoDL
2. apache-hive-3.1.2-bin.tar.gz
3. hadoop-2.10.0.tar.gz
4. postgresql-9.4.26-1-windows-x64.exe
5. sqoop-1.4.7.bin hadoop-2.6.0.tar.gz

BUKA TERMINAL

Install Java

1. Sudo apt update *(untuk merefresh system)*
2. Sudo apt install openjdk-8-jdk

```
tika@tika-VirtualBox:~$ sudo apt update
[sudo] password for tika:
Hit:1 http://id.archive.ubuntu.com/ubuntu focal InRelease
Hit:2 http://id.archive.ubuntu.com/ubuntu focal-updates InRelease
Hit:3 http://id.archive.ubuntu.com/ubuntu focal-backports InRelease
Hit:4 http://apt.postgresql.org/pub/repos/apt focal-pgdg InRelease
Hit:5 https://packages.microsoft.com/ubuntu/20.04/mssql-server-2019 focal InRel
ease
Hit:6 https://packages.microsoft.com/ubuntu/16.04/mssql-server-2019 xenial InRe
lease
Hit:7 https://packages.microsoft.com/ubuntu/20.04/prod focal InRelease
Hit:8 http://security.ubuntu.com/ubuntu focal-security InRelease
Reading package lists... Done
Building dependency tree
Reading state information... Done
570 packages can be upgraded. Run 'apt list --upgradable' to see them.
N: Skipping acquire of configured file 'main/binary-i386/Packages' as repositor
y 'http://apt.postgresql.org/pub/repos/apt focal-pgdg InRelease' doesn't suppor
t architecture 'i386'
tika@tika-VirtualBox:~$ sudo apt install openjdk-8-jdk
[sudo] password for tika:
Reading package lists... Done
Building dependency tree
Reading state information... Done
openjdk-8-jdk is already the newest version (8u312-b07-0ubuntu1~20.04).
0 upgraded, 0 newly installed, 0 to remove and 570 not upgraded.
tika@tika-VirtualBox:~$
```

Cek Versi Java

1. Java –version
2. sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
   export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
3. echo $JAVA_HOME

```
tika@tika-VirtualBox:~$ java -version
openjdk version "1.8.0_312"
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~20.04-b07)
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)
tika@tika-VirtualBox:~$ echo $JAVA_HOME
/usr/lib/jvm/java-8-openjdk-amd64
tika@tika-VirtualBox:~$
```

Install ssh

1. sudo apt install openssh-server openssh-client –y
2. cd ~/.ssh/
3. ssh-keygen -t rsa -P ""

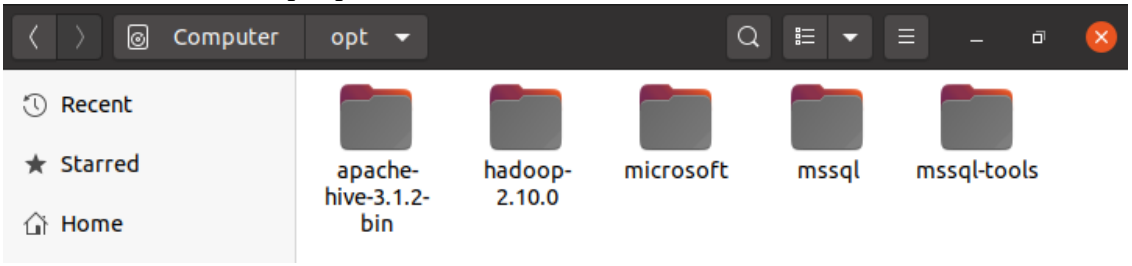
```
tika@tika-VirtualBox:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:8.2p1-4ubuntu0.4).
openssh-server is already the newest version (1:8.2p1-4ubuntu0.4).
0 upgraded, 0 newly installed, 0 to remove and 570 not upgraded.
tika@tika-VirtualBox:~$ cd ~/.ssh/
tika@tika-VirtualBox:~/.ssh$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/tika/.ssh/id_rsa):
/home/tika/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /home/tika/.ssh/id_rsa
Your public key has been saved in /home/tika/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:nAihNVc2lpp0g+bn9a3PpehzuSqvEmZclhpkISM7sM0 tika@tika-VirtualBox
The key's randomart image is:
+---[RSA 3072]----+
|   . = +o*o      |
|    B *+=*.      |
|   o E+ * . .    |
|      o+oo.=     |
|       .+S* . .  |
|         B   . . |
|        o .   . ..|
|          . . oooo |
+----[SHA256]-----+
```

4. cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
5. sudo /etc/init.d/ssh reload

```
tika@tika-VirtualBox:~/.ssh$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
tika@tika-VirtualBox:~/.ssh$ sudo /etc/init.d/ssh reload
Reloading ssh configuration (via systemctl): ssh.service.
```

Install Hive

1. sudo tar -xf ~/Downloads/apache-hive-3.1.2-bin.tar.gz /opt/
2. sudo chmod -R 777 /opt/apache-hive-3.1.2-bin/



3. hdfs dfs -mkdir /hive /hive/warehouse *(pilih salah 1)* hdfs dfs -chmod -R 775 /hive

4. sudo nano ~/.bashrc
   (Scrool sampai ke paling bawah dan ketikkan)
   export HIVE_HOME=/opt/apache-hive-3.1.2-bin
   export PATH=$PATH:$HIVE_HOME/bin
   export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.
   export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.

```
export HIVE_HOME=/opt/apache-hive-3.1.2-bin
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.
export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.
```

   (jika sudah klik ctrl+x, ketik y lalu enter)
5. sudo nano ${HIVE_HOME}/conf/hive-site.xml
   (Scrool sampai ke paling bawah dan ketikkan)
   ==========POSTGRES==========

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
    <name>hive.metastore.local</name>
    <value>true</value>
  </property>
  <property>
    <name>hive.metastore.warehouse.dir</name>
    <value>/hive/warehouse</value>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>org.postgresql.Driver</value>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:postgresql://localhost:5432/hivemetastoredb</value>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>root</value>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>root</value>
  </property>
  <property>
    <name>hive.server2.thrift.port</name>
    <value>10000</value>
```

```
    </property>
    <property>
       <name>hive.server2.enable.doAs</name>
       <value>true</value>
    </property>
    <property>
       <name>hive.execution.engine</name>
       <value>mr</value>
    </property>
    <property>
       <name>hive.metastore.port</name>
       <value>9083</value>
    </property>
    <property>
       <name>mapreduce.input.fileinputformat.input.dir.recursive</name>
       <value>true</value>
    </property>
</configuration>
==========================SQL SERVER==============================
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
       <name>hive.metastore.local</name>
       <value>true</value>
    </property>

    <property>
       <name>hive.metastore.warehouse.dir</name>
       <value>/hive/warehouse</value>
    </property>

    <property>
       <name>javax.jdo.option.ConnectionDriverName</name>
       <value>com.microsoft.sqlserver.jdbc.SQLServerDriver</value>
    </property>

    <property>
       <name>javax.jdo.option.ConnectionURL</name>
       <value>jdbc:sqlserver://localhost:1433;DatabaseName=hive_metastore;</value>
    </property>

    <property>
       <name>javax.jdo.option.ConnectionUserName</name>
       <value>sa</value>
    </property>
```

```
    <property>
      <name>javax.jdo.option.ConnectionPassword</name>
      <value>Admin123</value>
    </property>

    <property>
      <name>hive.server2.thrift.port</name>
      <value>10000</value>
    </property>

    <property>
      <name>hive.server2.enable.doAs</name>
      <value>true</value>
    </property>

    <property>
      <name>hive.execution.engine</name>
      <value>mr</value>
    </property>

    <property>
      <name>hive.metastore.port</name>
      <value>9083</value>
    </property>

    <property>
      <name>mapreduce.input.fileinputformat.input.dir.recursive</name>
      <value>true</value>
    </property>
<property>
  <name>hive.metastore.db.type</name>
  <value>mssql</value>
  <description>
    Expects one of [derby, oracle, mysql, mssql, postgres].
    Type of database used by the metastore. Information schema &amp;amp;
JDBCStorageHandler depend on it.
  </description>
</property>
</configuration>
```

(jika sudah klik ctrl+x, ketik y lalu enter)

6. create database hive_metastore di sql server

7. sudo nano ${HIVE_HOME}/bin/hive-config.sh
   (Scrool sampai ke paling bawah dan ketikkan)
   export HADOOP_HOME=/opt/hadoop-2.10.0
   (jika sudah klik ctrl+x, ketik y lalu enter)
8. ${HIVE_HOME}/bin/schematool -initSchema -dbType postgres
   ${HIVE_HOME}/bin/schematool -initSchema -dbType mssql
9. hiveserver2 & *(tunggu sampai 4 session)*

RUN HIVE di Mozilla
localhost:10002

Stop Hive

1. ps ax | grep hiveserver2
2. sudo kill -9 (id)

Masuk Db Hive

1. cd $HIVE_HOME/bin
2. hive
3. exit;

Install Hadoop

1. sudo tar -xf ~/Downloads/hadoop-2.10.0.tar.gz -C /opt/
2. sudo chmod -R 777 /opt/hadoop-2.10.0/
3. sudo nano ~/.bashrc
   (Scrool sampai ke paling bawah dan ketikkan)
   export JAVA_HOME=”/usr/lib/jvm/java-8-openjdk-amd64”
   export HADOOP_HOME=/opt/hadoop-2.10.0
   export HADOOP_INSTALL=$HADOOP_HOME
   export HADOOP_MAPRED_HOME=$HADOOP_HOME
   export HADOOP_COMMON_HOME=$HADOOP_HOME
   export HADOOP_HDFS_HOME=$HADOOP_HOME
   export YARN_HOME=$HADOOP_HOME
   export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
   export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
   export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
   export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:/opt/apache-hive-3.1.2-bin/lib/*

   export HIVE_HOME=/opt/apache-hive-3.1.2-bin
   export PATH=$PATH:$HIVE_HOME/bin
   export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.
   export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.

export PATH=”$PATH:/opt/mssql-tools/bin”

```
export JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64"
export HADOOP_HOME=/opt/hadoop-2.10.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:/opt/apache-hive-3.1.2-bin/lib/*
export HIVE_HOME=/opt/apache-hive-3.1.2-bin
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.
export CLASSPATH=$CLASSPATH:/opt/hadoop-2.10.0/lib/*:.

export PATH="$PATH:/opt/mssql-tools/bin"


^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
```

(jika sudah klik ctrl+x, ketik y lalu enter)

4. source ~/.bashrc
5. sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
   (Scrool sampai ke paling bawah dan ketikkan)
   export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

```
# NOTE: this should be set to a directory that can only be written to by
#       the user that will run the hadoop daemons.  Otherwise there is the
#       potential for a symlink attack.
export HADOOP_PID_DIR=${HADOOP_PID_DIR}
export HADOOP_SECURE_DN_PID_DIR=${HADOOP_PID_DIR}

# A string representing this instance of hadoop. $USER by default.
export HADOOP_IDENT_STRING=$USER
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

^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
```

(jika sudah klik ctrl+x, ketik y lalu enter)

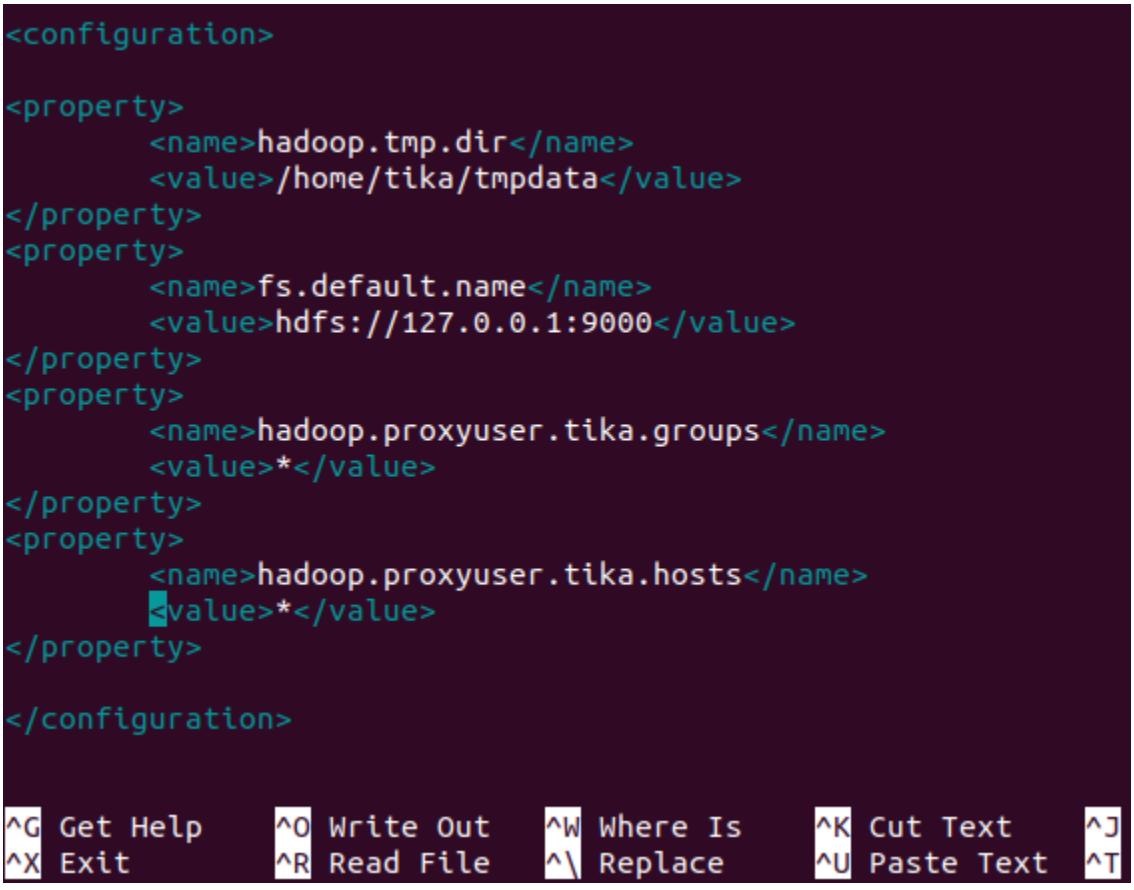
6. sudo nano ${HADOOP_HOME}/etc/hadoop/core-site.xml
   (Scrool sampai ke paling bawah dan ketikkan di dalam configuration)

```
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/tika/tmpdata</value>
</property>
```

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

<property>
      <name>hadoop.proxyuser.tika.groups</name>
      <value>*</value>
   </property>
   <property>
      <name>hadoop.proxyuser.tika.hosts</name>
      <value>*</value>
   </property>
```

----------jangan lupa create folder tmpdata----------



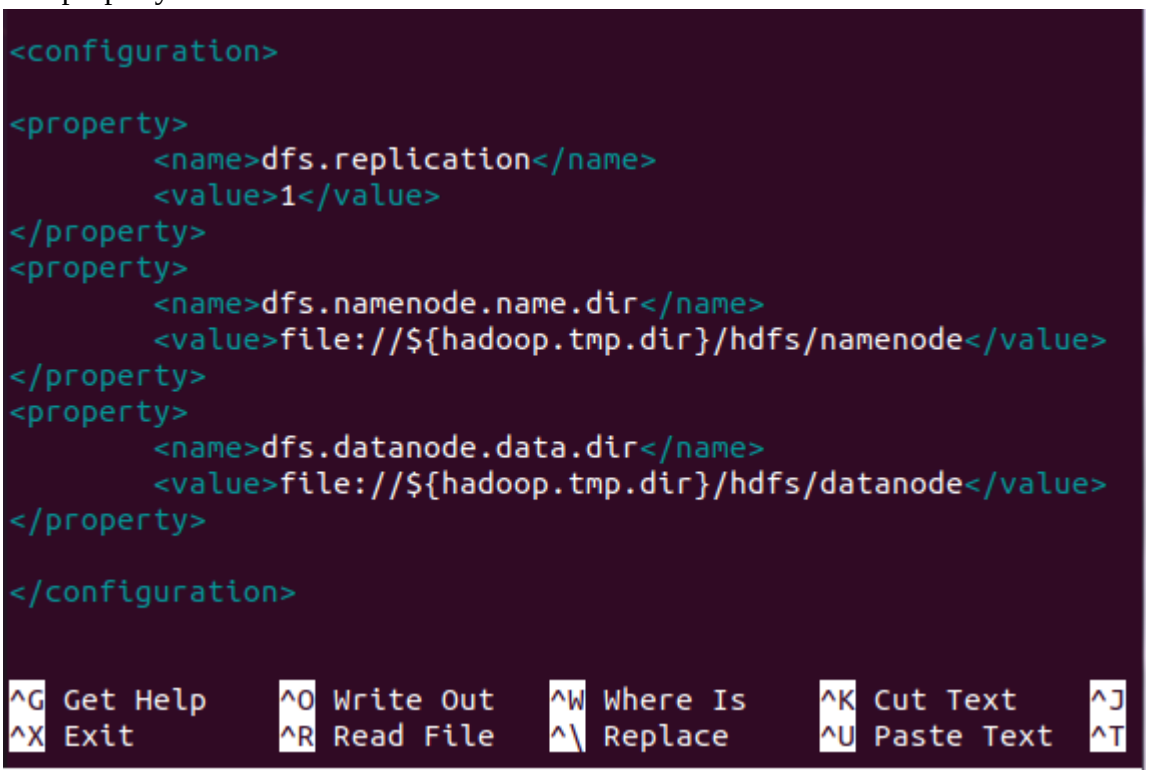
(jika sudah klik ctrl+x, ketik y lalu enter)

7. sudo nano ${HADOOP_HOME}/etc/hadoop/hdfs-site.xml
   (Scrool sampai ke paling bawah dan ketikkan di dalam configuration)

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

<property>
  <name>dfs.namenode.name.dir</name>
  <value>file://${hadoop.tmp.dir}/hdfs/namenode</value>
</property>

<property>
  <name>dfs.datanode.data.dir</name>
  <value>file://${hadoop.tmp.dir}/hdfs/datanode</value>
</property>
```



(jika sudah klik ctrl+x, ketik y lalu enter)

8. sudo nano ${HADOOP_HOME}/etc/hadoop/mapred-site.xml
   (Scrool sampai ke paling bawah dan ketikkan)

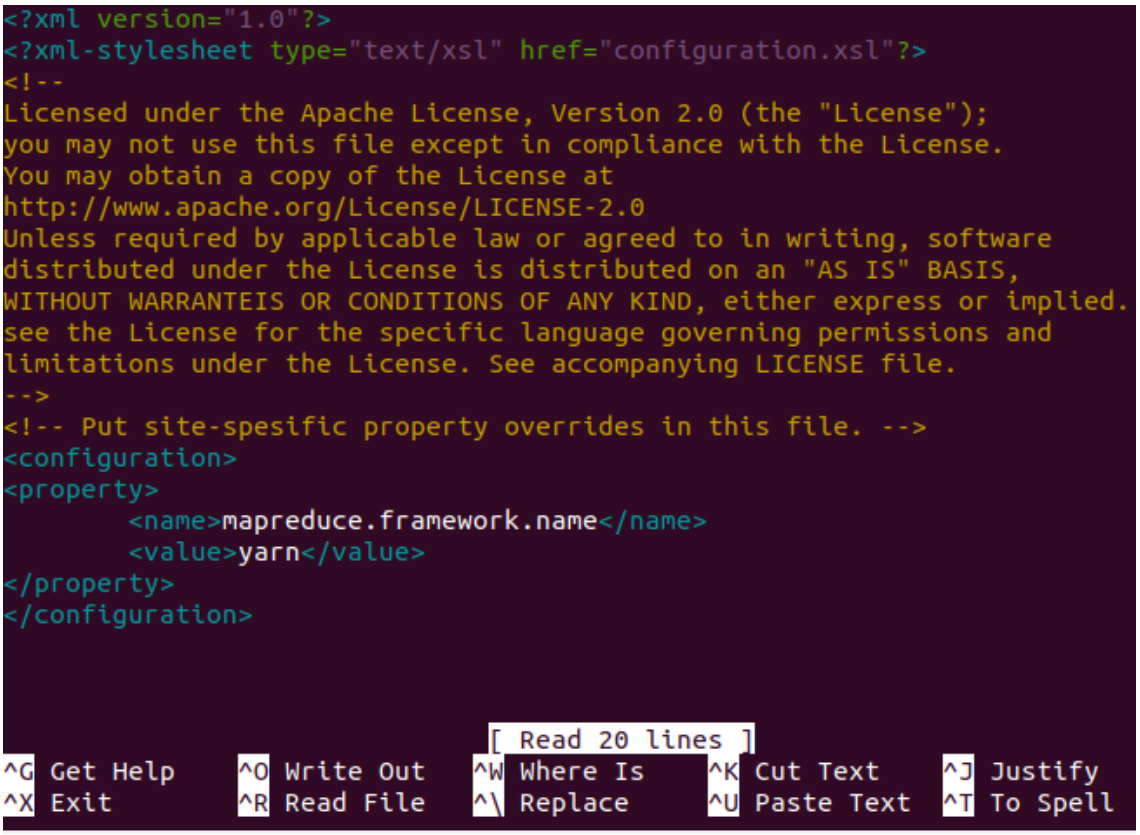
```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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>
```

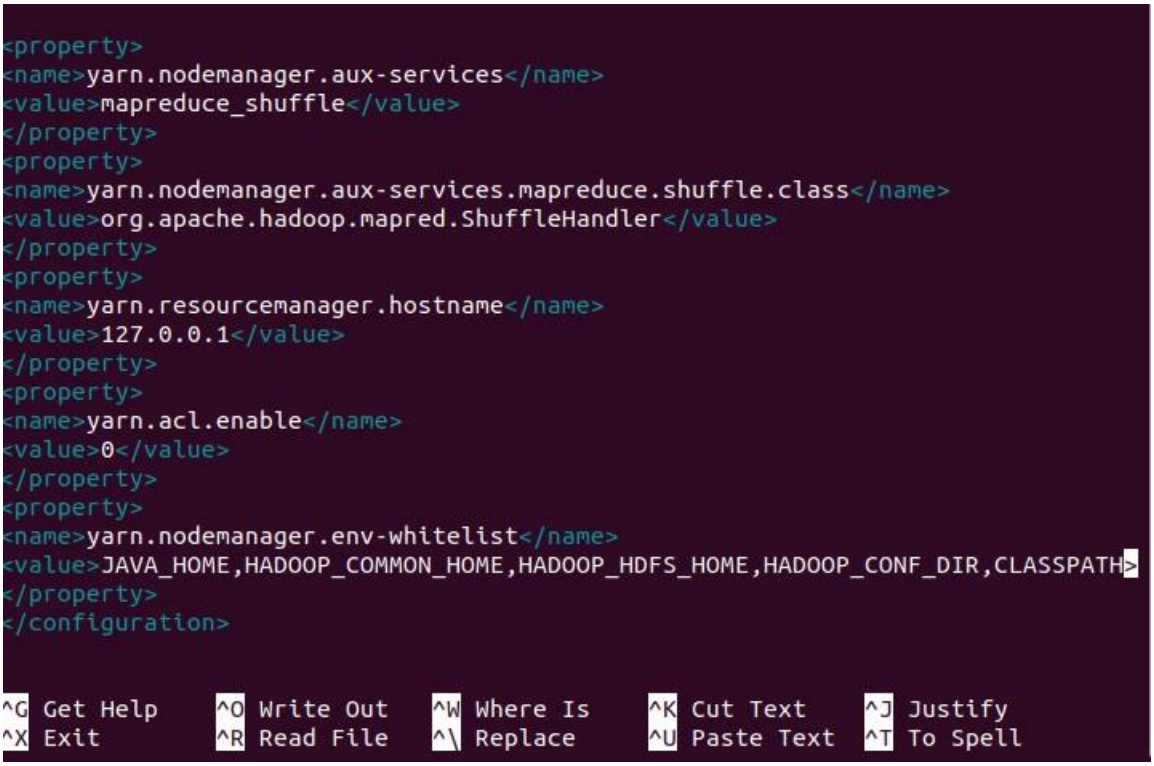


(jika sudah klik ctrl+x, ketik y lalu enter)

9. sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
   (Scrool sampai ke paling bawah dan ketikkan di dalam configutarion)

```
<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>



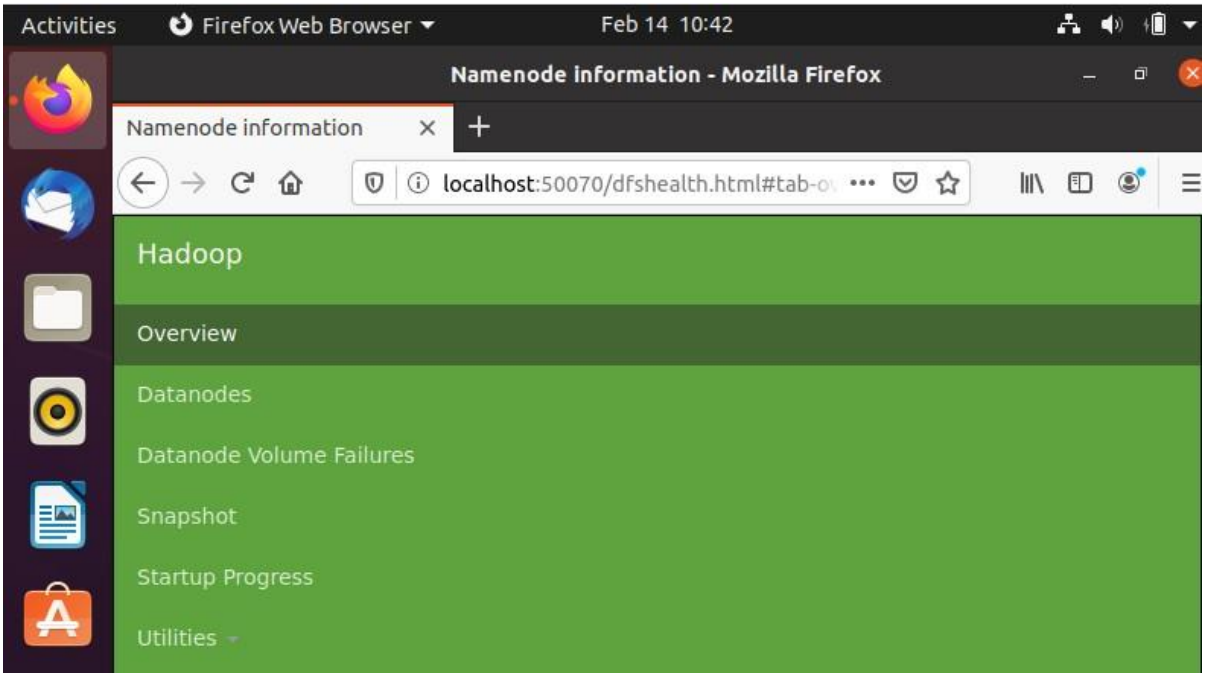
(jika sudah klik ctrl+x, ketik y lalu enter)

10. hadoop namenode –format
11. ${HADOOP_HOME}/sbin/start-all.sh *(menjalankan hadoop)*

```
tika@tika-VirtualBox:~/.ssh$ ${HADOOP_HOME}/sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hadoop-2.10.0/share/hadoop/common/lib/sl
f4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/apache-hive-3.1.2-bin/lib/log4j-slf4j-im
pl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation
.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Starting namenodes on [localhost]
localhost: starting namenode, logging to /opt/hadoop-2.10.0/logs/hadoop-tika-na
menode-tika-VirtualBox.out
localhost: starting datanode, logging to /opt/hadoop-2.10.0/logs/hadoop-tika-da
tanode-tika-VirtualBox.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /opt/hadoop-2.10.0/logs/hadoop-
tika-secondarynamenode-tika-VirtualBox.out
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hadoop-2.10.0/share/hadoop/common/lib/sl
f4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/apache-hive-3.1.2-bin/lib/log4j-slf4j-im
pl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation
.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
starting yarn daemons
```

RUN HADOOP di mozila

Localhost:50070/



Install Postgresql

3. Sudo apt update
4. Sudo apt install postgresql postgresql-contrib

```
tika@tika-VirtualBox:~$ sudo apt install postgresql postgresql-contrib
Reading package lists... Done
Building dependency tree
Reading state information... Done
Suggested packages:
  postgresql-doc
The following packages will be upgraded:
  postgresql postgresql-contrib
2 upgraded, 0 newly installed, 0 to remove and 568 not upgraded.
Need to get 133 kB of archives.
After this operation, 2.048 B of additional disk space will be used.
Get:1 http://apt.postgresql.org/pub/repos/apt focal-pgdg/main amd64 postgresql
all 14+238.pgdg20.04+1 [66,5 kB]
Get:2 http://apt.postgresql.org/pub/repos/apt focal-pgdg/main amd64 postgresql-
contrib all 14+238.pgdg20.04+1 [66,5 kB]
Fetched 133 kB in 2s (59,5 kB/s)
(Reading database ... 154633 files and directories currently installed.)
Preparing to unpack .../postgresql_14+238.pgdg20.04+1_all.deb ...
Unpacking postgresql (14+238.pgdg20.04+1) over (14+232.pgdg20.04+1) ...
Preparing to unpack .../postgresql-contrib_14+238.pgdg20.04+1_all.deb ...
Unpacking postgresql-contrib (14+238.pgdg20.04+1) over (14+232.pgdg20.04+1) ...
Setting up postgresql-contrib (14+238.pgdg20.04+1) ...
Setting up postgresql (14+238.pgdg20.04+1) ...
tika@tika-VirtualBox:~$
```

5. Sudo –i –u postgresql
6. Psql
7. Postgres=# \q
8. Exit

```
tika@tika-VirtualBox:~$ sudo -i -u postgres
postgres@tika-VirtualBox:~$ psql
psql (14.1 (Ubuntu 14.1-2.pgdg20.04+1), server 9.4.26)
Type "help" for help.

postgres=# \q
postgres@tika-VirtualBox:~$ exit
logout
```

9. Sudo –u postgres psql
10. Postgres=# \q
11. Sudo –u postgres createuser –interactive
    output => Enter name of role to add: tikalam
    Shall the new role be a superuser? (y/n) y
12. sudo -u postgres createdb tikalam
13. sudo adduser tikalam
14. sudo -i -u tikalam
15. psql
16. sudo su – postgres