NOTE :

- **text marked in grey are command to execute**
- **text marked in green are text to be modified in files**

1) first check whether java is install if yes then which version is installed

java -version

2) if not then run this command to install java

sudo apt-get install default-jre

3) after java install check for jps command working

jps

4) if not then run this command

sudo apt-get install openjdk-8-jdk

5) install ssh for remote login into data nodes

sudo apt-get install openssh-server

6) create a ssh public and private key

ssh-keygen

7) add the pub to key the authorised key file for password less access

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

8) try whether it works or not

ssh localhost

9) to get out use exit cmd

exit

10) now Untar the hadoop.tar.gz file tar -xvzf < path to hadoop downloaded file > hadoop-2.9.1.tar.gz

11) create a soft link for hadoop folder (optional**)**

ln -s hadoop-<Version of hadoop> hadoop

12) locate ur java path

readlink -f $(which java)

output will be /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java select only /usr/lib/jvm/java-8-openjdk-amd64

13) now open bashrc file and add Java and Hadoop Path

gedit ~/.bashrc

#inside this file at the last, add these line in green colour

export JAVA_HOME=< java path as seen in above point 12 >

export HADOOP_HOME=<path where hadoop downloaded file was extracted>

export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

example:-
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/cdac/Desktop/hadoop-3.2.0
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

#Save the file and source the file using

source ~/.bashrc

14) Now try hadoop version

hadoop version

output will the version of hadoop installed

15) Now again open the ~/.bashrc file and add the following lines

#NOTE ONLY ADD THOSE LINES WHICH ARE NOT THERE THE FINAL FILE SHOULD LOOK LIKE THE FOLLOWING (as marked in green)

gedit ~/.bashrc

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/cdac/Desktop/hadoop-3.2.0
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME export
HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin export
HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

16) Save the file and source the file using

source ~/.bashrc

17) NOW go to Hadoop folder

cd $HADOOP_HOME/etc/hadoop

17.1.)Open hadoop-env.sh file

gedit **hadoop-env.sh**
#at the last line of file add

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

#save the file

17.2)In the same directory i.e $HADOOP_HOME/etc/hadoop cp mapred-site.xml.template

open this file mapred-site.xml
gedit mapred-site.xml
#add following property line in between configuration line (marked in green)

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

#save the file

17.3.) similarly open core-site.xml
#add following property line in between configuration line (marked in green)

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

#save the file

17.4) now create to dir in your system where you want to store your data of namenode and datanode

1. **mkdir -p /home/cdac/Desktop/hdfs/namenode**
2. **mkdir -p /home/cdac/Desktop/hdfs/datanode**

gedit hdfs-site.xml

#add following property line in between configuration line (marked in green)

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>/home/cdac/Desktop/hdfs/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/home/cdac/Desktop/hdfs/datanode</value>

</property>
</configuration>
#save the file

17.5.) open this file yarn-site**.xml**

gedit yarn-site.xml
#add following property line in between configuration line (marked in green)

<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>

#save the file

18) Now format your namenode #NOTE this should be done only once
hadoop namenode -format

#output of this cmd should be succcesfull like this(marked in green)

18/06/23 12:09:36 INFO common.Storage: Storage directory /home/rohit/Desktop/hdfs/namenode has been successfully formatted.
18/06/23 12:09:36 INFO namenode.FSImageFormatProtobuf: Saving image file /home/rohit/Desktop/hdfs/namenode/current/fsimage.ckpt_0000000000000000000 using no compression
18/06/23 12:09:36 INFO namenode.FSImageFormatProtobuf: Image file /home/rohit/Desktop/hdfs/namenode/current/fsimage.ckpt_0000000000000000000 of size 322 bytes saved in 0 seconds .
18/06/23 12:09:36 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
18/06/23 12:09:36 INFO namenode.NameNode: SHUTDOWN_MSG:
/**
SHUTDOWN_MSG: Shutting down NameNode at rohit-Inspiron-5559/127.0.1.1
****

19) now start the hadoop by using

start-dfs.sh
start-yarn.sh

#after that check jps
jps

#it should display the following processes running

22672 ResourceManager
23088 Jps
22499 SecondaryNameNode
22804 NodeManager
22300 DataNode
22143 NameNode

Your hadoop is now up and runing you could check its ui at url http://localhost:50070

20) similarly to stop hadoop use stop-yarn.sh

Open the below links:

1.) **For NameNode:** http://localhost:50070

2.) **For Resource Manager:** http://localhost:8088

3.) **For Secondary NameNode:** http://localhost:50090

4.) **For DataNode:** http://localhost:5007

To Stop:

1.) **stop-dfs.sh**
2.) **stop-yarn.sh**