Program 4: Hadoop Installation

Following are installation commands for Hadoop on Ubuntu 18 / 20 / or ubuntu 22

The text written in comment (statements starting with #) are guideline instructions to read.
Read these commented instructions carefully

Open terminal and run following commands at terminal

(1.)

sudo apt update

2.)

sudo apt install openjdk-8-jdk

3.)

sudo nano .bashrc

Note : In this editor you can only use arrows keys
copy the text in following block BLOCK_1 and paste it at the end of above file
use command "shift + control + v" to copy paste ;
save file using "control o" () ;

exit using control x

#BLOCK_1 ++

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin
export HADOOP_HOME=~/hadoop-3.2.3/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh

++

4.)

sudo apt-get install ssh

5.) # Download hadoop-3.2.3 as follows

search "apache hadoop download" on google : follow the topmost link pointing to apache website
Here on this page somewhere you will find a link to archieves : click on the archieves link
On archives page It provides you a dircetory of hadoop installers of various version
choose directory "hadoop-3.2.3"
Here download the following file : hadoop-3.2.3.tar.gz
let the download finish ; then extract the file using following command

tar -zxvf ~/Downloads/hadoop-3.2.3.tar.gz

Note: instead of tilde sign in above , you may give full path of the downloaded installer

6.) # change the directory as follows

cd hadoop-3.2.3/etc/hadoop

7) # open hadoop-env.sh using following command

sudo nano hadoop-env.sh

here in this file you will find the line like ..... JAVA_HOME=/usr/java/testing hdfsdfs -ls
go to this line and remove the comment of this line also modify this line as shown below
note after removing comment the line should not have any leading spaces

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

Additionaly one more line in above file has JAVA_Home: Do the similar modifications there

save this file after above two modifications and exit
8.) # open file core-site.xml as follows

sudo nano core-site.xml

copy the text in the following block BLOCK2 and paste it in above file in between <configuration>
tags

#BLOCK2 ######################################

<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value> </property>
<property>
<name>hadoop.proxyuser.dataflair.groups</name> <value>*</value>
</property>
<property>
<name>hadoop.proxyuser.dataflair.hosts</name> <value>*</value>
</property>
<property>
<name>hadoop.proxyuser.server.hosts</name> <value>*</value>
</property>
<property>

<name>hadoop.proxyuser.server.groups</name> <value>*</value>
</property>
</configuration>
##

9.) # open file hdfs-site.xml as follows

sudo nano hdfs-site.xml
# copy the text in the following block BLOCK3 and paste it in above file in between <configuration>
# tags

BLOCK3**********

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

10.) # open the file mapred-site.xml as follows

sudo nano mapred-site.xml
# copy the text in the following block BLOCK4 and paste it in above file in between <configuration>
# tags

BLOCK4 %%

 <property>
 <name>mapreduce.framework.name</name> <value>yarn</value>
 </property>
 <property>
 <name>mapreduce.application.classpath</name>

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
 </property>

%%

11.) # open the file yarn-site.xml as follows

sudo nano yarn-site.xml
# copy the text in the following block BLOCK5 and paste it in above file in between <configuration>
# tags

BLOCK 5===

 <property>

<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
==========

12.)

connecting to ssh ; here give your Systems (OS user) password when asked

ssh localhost

(13.)

sudo service ssh restart

14.)

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

15.)

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

16.)

chmod 0600 ~/.ssh/authorized_keys

17.)

hadoop-3.2.3/bin/hdfs namenode -format

18.)

#format the HDFS file system using following command

export PDSH_RCMD_TYPE=ssh

19.)

start hadoop as follows

start-all.sh

20.)

check your Hadoop installtion by entering following command

jps

It should show all the following dameons running

- NameNode
- DataNode
- Secondary Name Node
- Resource Manager
- Node Manager

21.)

OR goto browser tpye the address : "localhost:9870" to check HDFS directory

22.)

Or you may test your HDFS by creating directories and files by running follwing commands at
terminal

```
hadoop fs -mkdir /user
hadoop fs -mkdir /user/MNK
hadoop fs -put demo.csv /user/MNK
```