Pre requisite: - virtual box and ubuntu

step1: download java

write this command on terminal:

sudo apt-get install openjdk-8-jre-headless

to check if it is downloaded or not: java -version

path to check java availability: file - other location - computer

usr/lib/jvm

step2: copy paste the link in browser of ubuntu https://hadoop.apache.org/releases.html

version to download - 3.2.4 - binary

note - a tar.gz file will be downloaded in download folder

cut it from download folder and paste it on home folder after that extract it in home folder.

step 3: terminal command

sudo apt-get install ssh

step 4: terminal command

ssh-keygen -t rsa -P ""

step5: terminal command

cat /home/shivani/.ssh/id\_rsa.pub >> /home/shivani/.ssh/authorized\_keys

step6: terminal command

gedit ~/.bashrc

a new file will open with this command

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP\_INSTALL=/home/shivani/hadoop-3.2.4

export PATH=$PATH:$HADOOP\_INSTALL/bin

export PATH=$PATH:$HADOOP\_INSTALL/sbin6.cat /home/shivani/.ssh/id\_rsa.pub >> /home/shivani/.ssh/authorized\_keys

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

export YARN\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

step 7: go to this particular location - files/home/hadoop3.2.4/hadoop/etc/hadoop-env.sh

make this particular change on line no 36 after removing # symbol

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

step 8: terminal command - sudo mkdir -p /home/shivani/hadoop-3.2.4/tmp

go to this particular location - files/home/hadoop3.2.4/hadoop/etc/core-site.xml

<property>

<name>hadoop.tmp.dir</name>

<value>/home/shivani/hadoop-3.2.4/tmp</value>

<description>A base for other temporary directories.</description>

</property>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:54310</value>

<description>The name of the default file system. A URI whose

scheme and authority determine the FileSystem implementation. The

uri's scheme determines the config property (fs.SCHEME.impl) naming

the FileSystem implementation class. The uri's authority is used to

determine the host, port, etc. for a filesystem.</description>

</property>

Step9: go to this particular location - files/home/hadoop3.2.4/hadoop/etc/ mapred-site.xml

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>localhost:54311</value>

<description>The host and port that the MapReduce job tracker runs

at. If "local", then jobs are run in-process as a single map

and reduce task.

</description>

</property>

</configuration>

Step10: command on terminal

sudo mkdir -p /usr/local/hadoop\_store/hdfs/namenode

sudo mkdir -p /usr/local/hadoop\_store/hdfs/datanode

go to this particular location - files/home/hadoop3.2.4/hadoop/etc/hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>file:/usr/local/hadoop\_store/hdfs/namenode</value>

</property>

<property>

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

<value>file:/usr/local/hadoop\_store/hdfs/datanode</value>

</property>

Step11:terminal command

hadoop namenode -format

start-all.sh

jps