Hadoop Installation Steps

1. Update Ubuntu

```
sudo apt-get update
```

2. Make sure java is installed

```
java --version
```

if java is not installed, then type

```
sudo apt-get install default-jdk
```

3. Add group hadoop

```
sudo addgroup hadoop
```

4. Make hduser a super user (Administrator)

```
sudo adduser hduser sudo
```

5. Install ssh server

```
sudo apt-get install ssh-server
```

Generate public/private RSA key pair

```
ssh-keygen -t rsa -P ""
```

When prompted for the file name to save the key, press enter (leave it blank)

Type the following commands

```
-> cat $HOME /.ssh/id_rsa.pub >> $HOME /.ssh/authorized_keys
-> ssh localhost
-> exit
```

6. Install Hadoop

Google for

```
apache hadoop download index 3.3.4
```

Download hadoop_3.3.4.tar.gz
or
go to
hadoop-3.3.4.tar.gz

7. Unzip Hadoop file

Once downloaded, open the terminal and cd to directory where it is downloaded and extract it as follows:-

```
-> cd Downloads
-> sudo tar -xvzf hadoop-3.3.4.tar.gz
```

You can now check that there is an extra file named “hadoop-3.3.4” by typing the command “ls”.

8. Move the hadoop file

```
sudo mv hadoop-3.3.4 /usr/local/hadoop
```

9. Make hduser the owner of /usr/local

```
sudo chown -R hduser /usr/local
```

10. Configure hadoop system

open ~/.bashrc

```
sudo gedit ~/.bashrc
```

At the end of the file, add the following lines

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
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=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
```

Now source ~/.bashrc

```
source ~/.bashrc
```

Now open hadoop-env.sh

```
sudo gedit /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

search for the line starting with “export JAVA_HOME=” and replace it with the following line,

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

save the file

-> open core-site.xml

```
sudo gedit /usr/local/hadoop/etc/hadoop/core-site.xml
```

Add the following lien between the tags

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

-> Open hdfs-site.xml

```
 sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

Add the following lines between the tags

```
<property>
      <name>dfs.replication</name>
      <value>1</value>
</property>
<property>
      <name>dfs.namenode.name.dir</name>
      <value>file:/usr/local/hadoop_tmp/hdfs/namenode</value>
</property>
<property>
     <name>dfs.datanode.data.dir</name>
     <value>file:/usr/local/hadoop_tmp/hdfs/datanode</value>
</property>
```

->Open yarn-site.xml

```
sudo gedit /usr/local/hadoop/etc/hadoop/yarn-site.xml
```

Add the following lines between the tags

```
<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>


 OR


<property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
    <name>yarn.nodemanager.resource.memory-mb</name>
    <value>40960</value>
</property>
<property>
 <name>yarn.scheduler.minimum-allocation-mb</name>
 <value>2048</value>
</property>
```

-> Open mapred-site.xml

```
sudo gedit /usr/local/hadoop/etc/hadoop/mapred-site.xml
```

Add the following lines between the lines,

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

 OR

<property>
  <name>mapreduce.framework.name</name>
  <name>mapreduce.map.memory.mb</name>

 <value>4096</value>

 <name>mapreduce.reduce.memory.mb</name>

 <value>8192</value>
 <name>mapreduce.map.java.opts</name>

 <value>-Xmx3072m</value>

 <name>mapreduce.reduce.java.opts</name>
```

```
 <value>-Xmx6144m</value>
</property>
```

11. Make datanode and namenode

Now run the following commands on the terminal to create a directory for hadoop-tmp namenode and datanode

```
sudo mkdir -p /usr/local/hadoop-tmp
sudo mkdir -p /usr/local/hadoop-tmp/namenode
sudo mkdir -p /usr/local/hadoop-tmp/datanode
```

12. make hduser owner of /usr/local

```
sudo chown -R hduser /usr/local
```

13. Format the namenode as follows

```
hdfs namenode -format
```

14. Start the HDFS file system

```
start-dfs.sh
```

15. Start the yarn

```
start-yarn.sh
```

16. Type the following command

```
jps
```

make sure these node are listed

```
Resource Manager
NameNode
Node Manager
Secondary NameNode
Jps
DataNode
```

17. Go to localhost:9870 or localhost:50070

you will see hadoop UI