Step 1 : Install java jdk 8

First of all you must install Java JDK 8 on your system. You can just type this command to install java jdk on your system.

sudo apt install openjdk-8-jdk

To check it's there **cd /usr/lib/jvm**

Step 2 : Add this configuration on you bash file

Now just open .bashrc file and paste these commands.

sudo nano .bashrc

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-11-openjdk-amd64/bin
export HADOOP_HOME=~/hadoop-3.4.1
export PATH=$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
```

(ssh — secure shell — protocol used to securely connect to remote server/system — transfers data in encrypted form)

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

now go to hadoop.apache.org website download the tar file
(hadoop.apache.org — download tar file of hadoop.)

```
tar -zxvf hadoop-3.4.1.tar.gz
```

(Extract the tar file)
cd hadoop-3.4.1/etc/hadoop

now open hadoop-env.h

sudo nano hadoop-env.sh

```
JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```
(set the path for JAVA_HOME)

Step 3 : Add this file in core-site.xml

Now add this configuration in core-site.xml file.

sudo nano core-site.xml

```
<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>
```

Step 3 : Add this file in hdfs-site.xml

Now add this configuration in hdfs-site.xml file.

sudo nano hdfs-site.xml

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

Step 4: Add this file in mapred-site.xml

Now add this configuration in mapred-site.xml file.

sudo nano mapred-site.xml

```
<configuration>
 <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>
</configuration>
```

Step 4: Add this file in yarn-site.xml

Now add this configuration in yarn-site.xml file.

sudo nano yarn-site.xml

```
<configuration>
 <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_PREP END_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
 </property>
</configuration>
```

ssh

```
ssh localhost

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

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

chmod 0600 ~/.ssh/authorized_keys

hadoop-3.4.1/bin/hdfs namenode -format
```

format the file system

```
export PDSH_RCMD_TYPE=ssh
```

Step 5 : Start hadoop

To start

start-all.sh **(Start NameNode daemon and DataNode daemon)**

localhost:9870

localhost:8088

codewitharjun@cwa:~$ hadoop fs -mkdir /user

codewitharjun@cwa:~$ hadoop fs -mkdir /user/arjun.gautam

codewitharjun@cwa:~$ touch demo.csv

codewitharjun@cwa:~$ hadoop fs -put demo.csv /user/arjun.gautam