

Step 1 : Install java jdk 8

sudo apt install openjdk-8-jdk

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

Step 2 : Add this configuration on bash file

nano ~/.bashrc

open .bashrc file and paste these commands.

```
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
```

Step 3: Execute bashrc file

source ~/.bashrc

Install ssh

sudo apt-get install ssh

Step 4: hadoop.apache.org website to download the tar file

Extract the tar file

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

Step 5: Add the java path in hadoop-env.h

cd hadoop-3.2.3/etc/hadoop

nano hadoop-env.h

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

Step 6: Add this file in core-site.xml

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 7 : Add this file in hdfs-site.xml

hdfs-site.xml

```

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

```

Step 8: Add this file in mapred-site.xml

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 9: Add this file in yarn-site.xml

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,CLAS
SPATH_PREP END_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

Step 10: 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
```

step 11: namenode -format

```
cd /hadoop-3.2.3/bin
hdfs namenode -format
```

format the file system

```
export PDSH_RCMD_TYPE=ssh
```

Step 12 : Start hadoop

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

Step 13: open localhost:9870 in browser.