Roll No: 20BCE204

Course Name and Course Code: 2CS702 Big Data Analytics

Practical No: 3

Aim: Install and configure single node Hadoop cluster. Perform HDFS commands on singlenode Hadoop Cluster.

Steps of Installation:

**Install Java JDK and JAVA HOME**

/usr/libexec/java\_home

echo $JAVA\_HOME

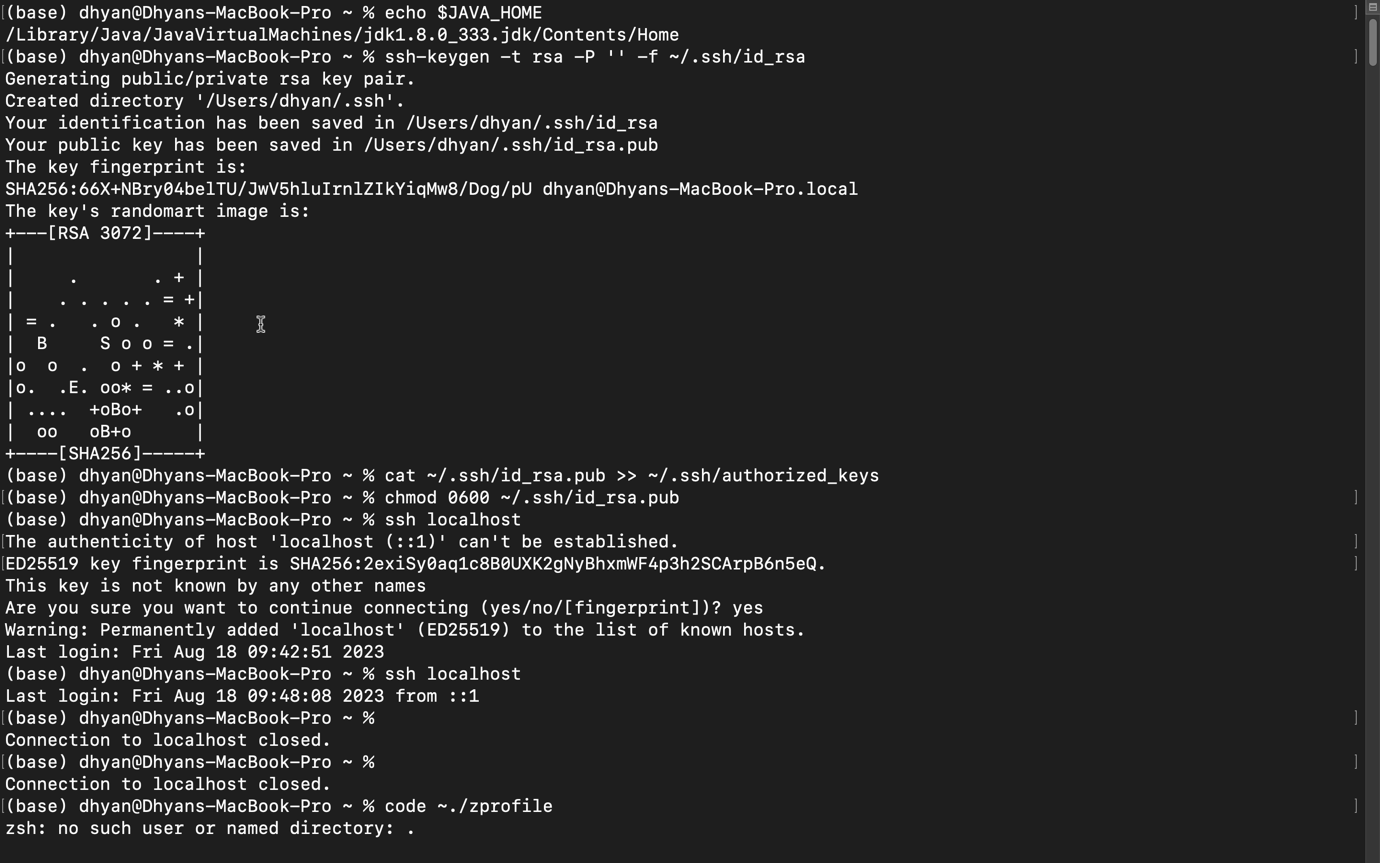
**Enable SSH to localhost**

ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

chmod 0600 ~/.ssh/id\_rsa.pub

ssh localhost



**Then download hadoop :**

once you download and extract set up the environment variables by opening .zprofile from root directory.

# Hadoop   
export HADOOP\_HOME=/Users/dhyan /hadoop-3.3.6/   
export HADOOP\_INSTALL=$HADOOP\_HOME   
export HADOOP\_MAPRED\_HOME=$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 PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin   
export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib/nativ"

**Now Configure Hadoop**

sudo code $HADOOP\_HOME/etc/hadoop/hadoop-env.sh

Then edit core-site.xml

<configuration>

<property>

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

<value>/Users/dhyan/hdfs/tmp/</value>

</property>

<property>

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

<value>hdfs://127.0.0.1:9000</value>

</property>

</configuration>

Edit hdfs-site.xml

configuration>

<property>

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

<value>/Users/dhyan/hdfs/namenode</value>

</property>

<property>

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

<value>/Users/dhyan/hdfs/datanode</value>

</property>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

Edit mapred-site.xml

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

Edit yarn-site.xml

<configuration>

<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>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>127.0.0.1</value>

</property>

<property>

<name>yarn.acl.enable</name>

<value>0</value>

</property>

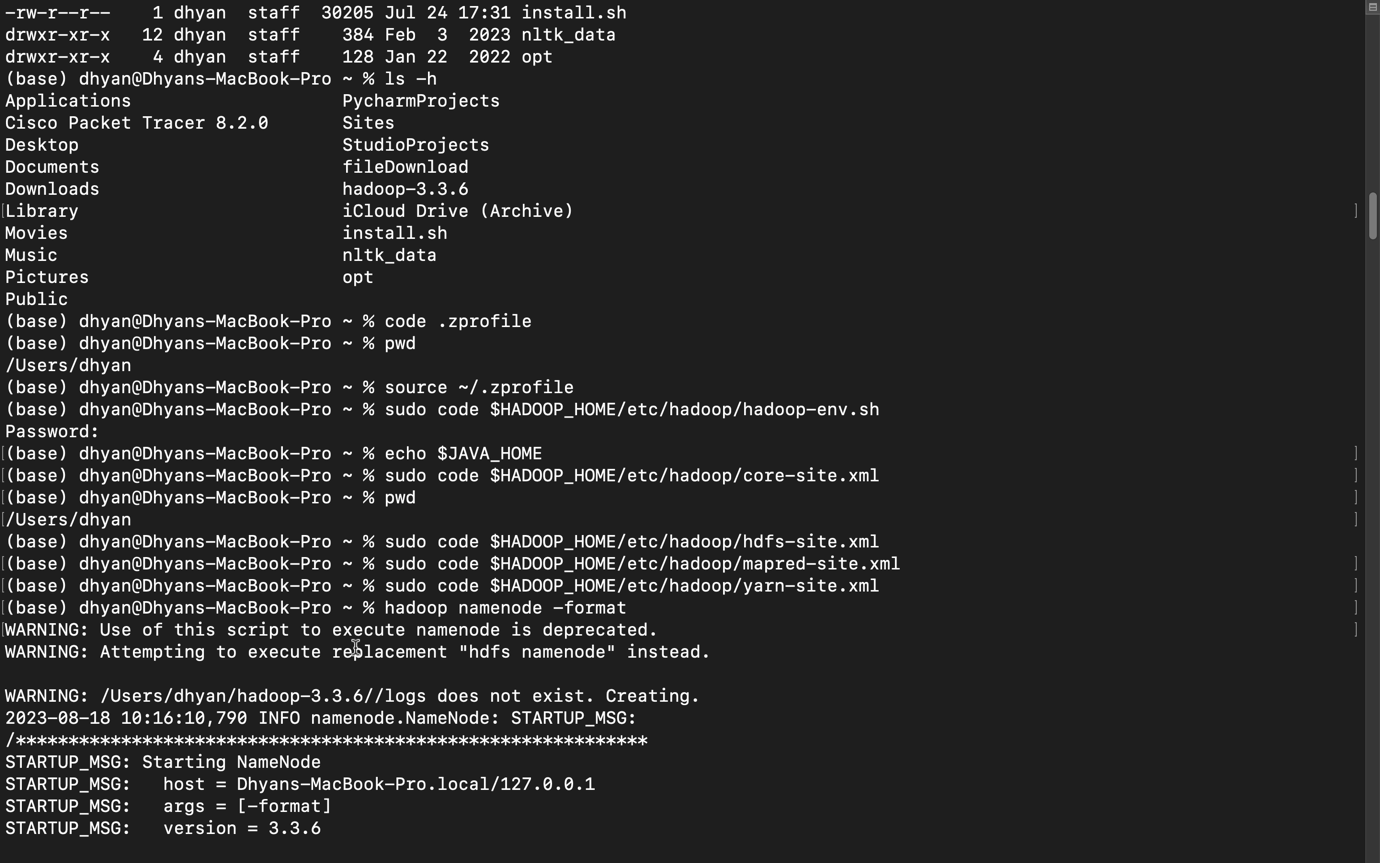
<property>

<name>yarn.nodemanager.env-whitelist</name>

<value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PERPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME</value>

</property>

</configuration>

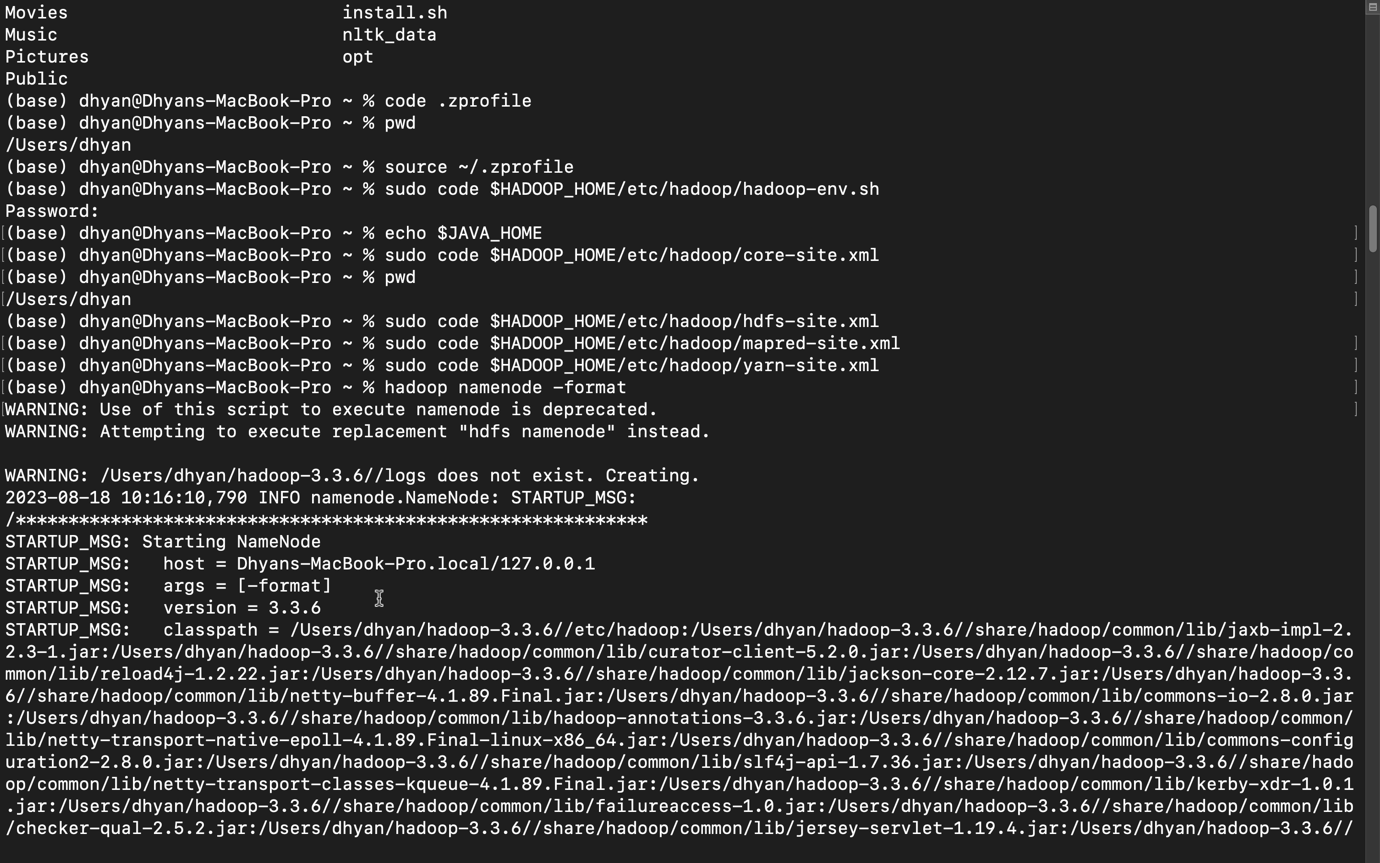


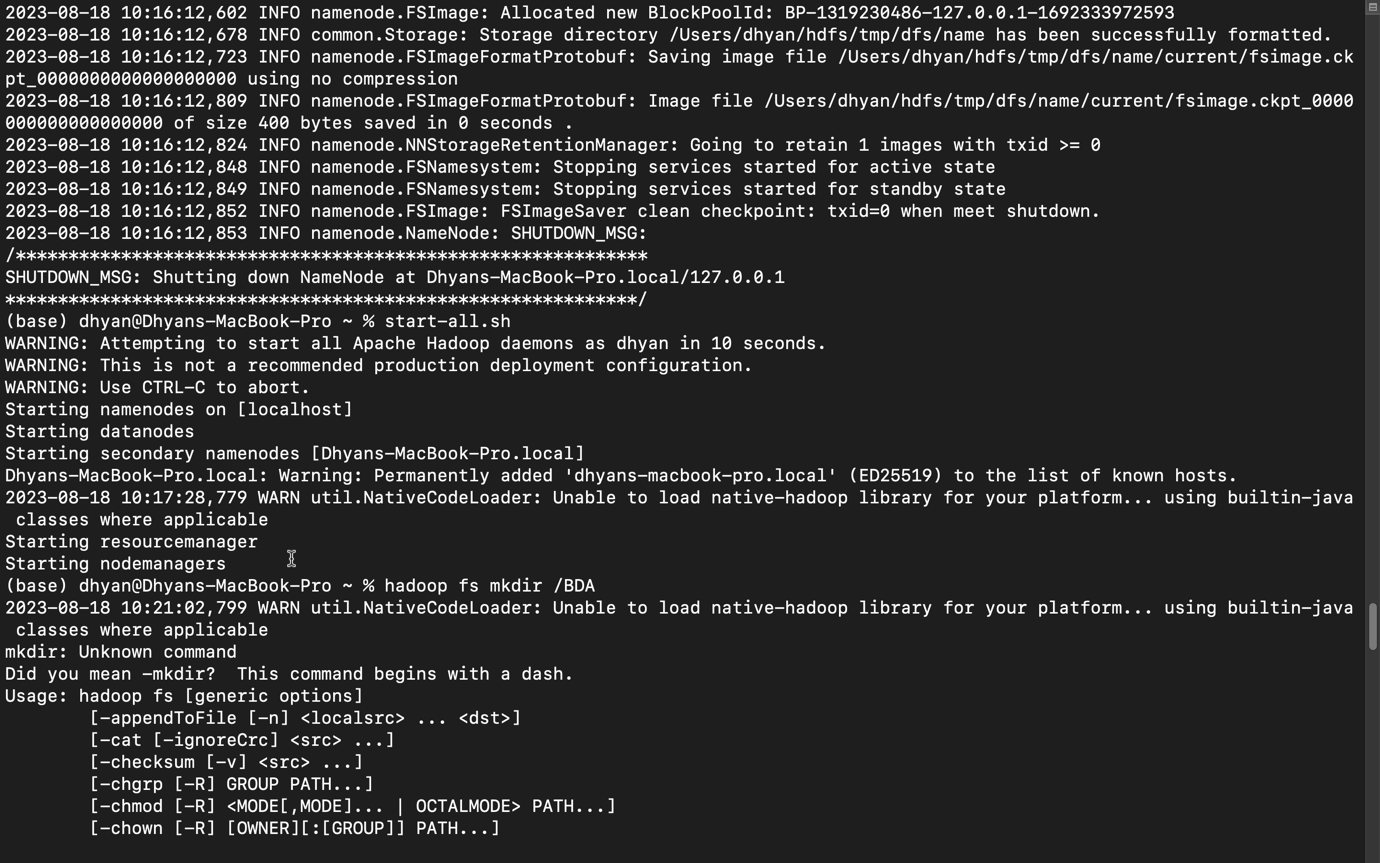
**Now Format the HDFS namenode.**

hdfs namenode -format

**Then start all of Hadoop :**

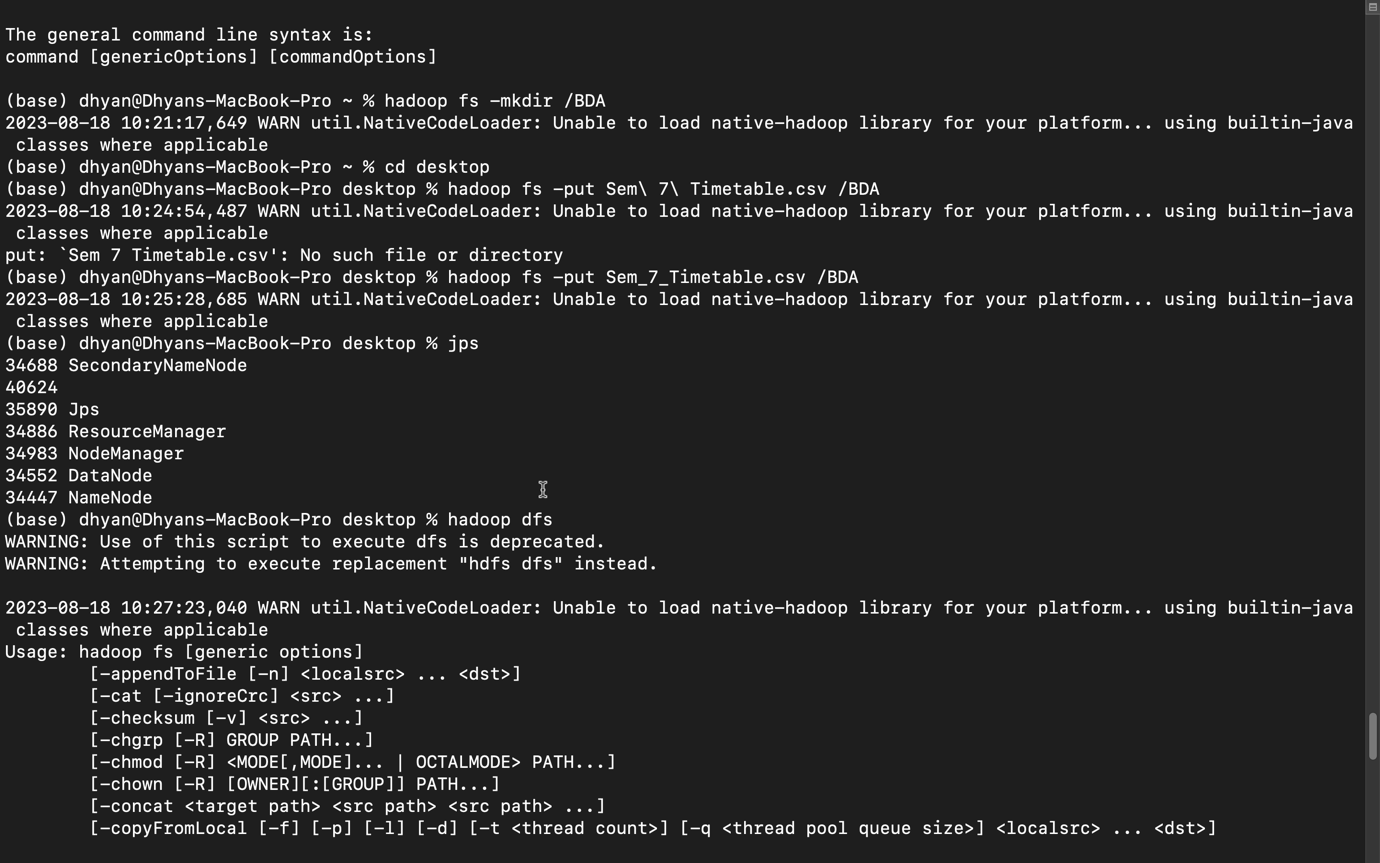
start-all.sh





**Creating new folder**

hadoop fs -mkdir /BDA



**List the content of a directory**

hadoop fs -ls /BDA

**Upload and download a file in HDFS**

hadoop fs -get /BDA/Sem\_7\_Timetable.csv /Users

hadoop fs -put Sem\_7\_Timetable.csv /BDA

**Look at the content of a file**

hadoop fs -cat /BDA/Sem\_7\_Timetable.csv

