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

```
(base) dhyan@Dhyans-MacBook-Pro ~ % echo $JAVA_HOME
/Library/Java/JavaVirtualMachines/jdk1.8.0_333.jdk/Contents/Home
(base) dhyan@Dhyans-MacBook-Pro ~ % ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/Users/dhyan/.ssh'.
Your identification has been saved in /Users/dhyan/.ssh/id_rsa
Your public key has been saved in /Users/dhyan/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:66X+NBry04belTU/JwV5hluIrnlZIkYiqMw8/Dog/pU dhyan@Dhyans-MacBook-Pro.local
The key's randomart image is:
+---[RSA 3072]----+
|                 |
|    .       . +  |
|   . . . . . = + |
| = .   . o .   * |
|  B     S o o = .|
|o  o  .  o + * + |
|o.  .E. oo* = ..o|
| ....  +oBo+   .o|
|  oo   oB+o       |
+----[SHA256]-----+
(base) dhyan@Dhyans-MacBook-Pro ~ % cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
(base) dhyan@Dhyans-MacBook-Pro ~ % chmod 0600 ~/.ssh/id_rsa.pub
(base) dhyan@Dhyans-MacBook-Pro ~ % ssh localhost
The authenticity of host 'localhost (::1)' can't be established.
ED25519 key fingerprint is SHA256:2exiSy0aq1c8B0UXK2gNyBhxmWF4p3h2SCArpB6n5eQ.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Last login: Fri Aug 18 09:42:51 2023
(base) dhyan@Dhyans-MacBook-Pro ~ % ssh localhost
Last login: Fri Aug 18 09:48:08 2023 from ::1
(base) dhyan@Dhyans-MacBook-Pro ~ %
Connection to localhost closed.
(base) dhyan@Dhyans-MacBook-Pro ~ %
Connection to localhost closed.
(base) dhyan@Dhyans-MacBook-Pro ~ % code ~./zprofile
zsh: no such user or named directory: .
```

Then download hadoop :

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

Hadoop
export HADOOP_HOME=/Users/arjuncodes/hadoop-3.3.1/
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>
```

```xml
    <value>1</value>
  </property>
</configuration>
```

Edit mapred-site.xml

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


</configuration>
```

Edit yarn-site.xml

```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>

```
(base) dhyan@Dhyans-MacBook-Pro ~ % code .zprofile
(base) dhyan@Dhyans-MacBook-Pro ~ % pwd
/Users/dhyan
(base) dhyan@Dhyans-MacBook-Pro ~ % source ~/.zprofile
(base) dhyan@Dhyans-MacBook-Pro ~ % sudo code $HADOOP_HOME/etc/hadoop/hadoop-env.sh
Password:
(base) dhyan@Dhyans-MacBook-Pro ~ % echo $JAVA_HOME
(base) dhyan@Dhyans-MacBook-Pro ~ % sudo code $HADOOP_HOME/etc/hadoop/core-site.xml
(base) dhyan@Dhyans-MacBook-Pro ~ % pwd
/Users/dhyan
(base) dhyan@Dhyans-MacBook-Pro ~ % sudo code $HADOOP_HOME/etc/hadoop/hdfs-site.xml
(base) dhyan@Dhyans-MacBook-Pro ~ % sudo code $HADOOP_HOME/etc/hadoop/mapred-site.xml
(base) dhyan@Dhyans-MacBook-Pro ~ % sudo code $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

Now Format the HDFS namenode.

hdfs namenode -format

Then start all of Hadoop :

start-all.sh

```
(base) dhyan@Dhyans-MacBook-Pro ~ % hadoop namenode -format
WARNING: Use of this script to execute namenode is deprecated.
WARNING: Attempting to execute replacement "hdfs namenode" instead.

WARNING: /Users/dhyan/hadoop-3.3.6//logs does not exist. Creating.
2023-08-18 10:16:10,790 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = Dhyans-MacBook-Pro.local/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = /Users/dhyan/hadoop-3.3.6//etc/hadoop:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/jaxb-impl-2.
2.3-1.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/curator-client-5.2.0.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/co
mmon/lib/reload4j-1.2.22.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/jackson-core-2.12.7.jar:/Users/dhyan/hadoop-3.3.
6//share/hadoop/common/lib/netty-buffer-4.1.89.Final.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/commons-io-2.8.0.jar
:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/hadoop-annotations-3.3.6.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/
lib/netty-transport-native-epoll-4.1.89.Final-linux-x86_64.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/commons-config
uration2-2.8.0.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/slf4j-api-1.7.36.jar:/Users/dhyan/hadoop-3.3.6//share/hado
op/common/lib/netty-transport-classes-kqueue-4.1.89.Final.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/kerby-xdr-1.0.1
.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/failureaccess-1.0.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib
/checker-qual-2.5.2.jar:/Users/dhyan/hadoop-3.3.6//share/hadoop/common/lib/jersey-servlet-1.19.4.jar:/Users/dhyan/hadoop-3.3.6//
```

```
2023-08-18 10:16:12,602 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1319230486-127.0.0.1-1692333972593
2023-08-18 10:16:12,678 INFO common.Storage: Storage directory /Users/dhyan/hdfs/tmp/dfs/name has been successfully formatted.
2023-08-18 10:16:12,723 INFO namenode.FSImageFormatProtobuf: Saving image file /Users/dhyan/hdfs/tmp/dfs/name/current/fsimage.ck
pt_0000000000000000000 using no compression
2023-08-18 10:16:12,809 INFO namenode.FSImageFormatProtobuf: Image file /Users/dhyan/hdfs/tmp/dfs/name/current/fsimage.ckpt_0000
000000000000000 of size 400 bytes saved in 0 seconds .
2023-08-18 10:16:12,824 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2023-08-18 10:16:12,848 INFO namenode.FSNamesystem: Stopping services started for active state
2023-08-18 10:16:12,849 INFO namenode.FSNamesystem: Stopping services started for standby state
2023-08-18 10:16:12,852 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2023-08-18 10:16:12,853 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at Dhyans-MacBook-Pro.local/127.0.0.1
************************************************************/
(base) dhyan@Dhyans-MacBook-Pro ~ % start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as dhyan in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Dhyans-MacBook-Pro.local]
Dhyans-MacBook-Pro.local: Warning: Permanently added 'dhyans-macbook-pro.local' (ED25519) to the list of known hosts.
2023-08-18 10:17:28,779 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
Starting resourcemanager
Starting nodemanagers
```

Creating new folder

hadoop fs -mkdir /BDA

```
The general command line syntax is:
command [genericOptions] [commandOptions]

(base) dhyan@Dhyans-MacBook-Pro ~ % hadoop fs -mkdir /BDA
2023-08-18 10:21:17,649 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
(base) dhyan@Dhyans-MacBook-Pro ~ % cd desktop
(base) dhyan@Dhyans-MacBook-Pro desktop % hadoop fs -put Sem\ 7\ Timetable.csv /BDA
2023-08-18 10:24:54,487 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
put: `Sem 7 Timetable.csv': No such file or directory
(base) dhyan@Dhyans-MacBook-Pro desktop % hadoop fs -put Sem_7_Timetable.csv /BDA
2023-08-18 10:25:28,685 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
(base) dhyan@Dhyans-MacBook-Pro desktop % jps
34688 SecondaryNameNode
40624
35890 Jps
34886 ResourceManager
34983 NodeManager
34552 DataNode
34447 NameNode
```

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

```
 classes where applicable
(base) dhyan@Dhyans-MacBook-Pro / % hadoop fs -ls /BDA
2023-08-19 15:38:33,475 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
Found 1 items
-rw-r--r--   1 dhyan supergroup       1109 2023-08-18 10:25 /BDA/Sem_7_Timetable.csv
(base) dhyan@Dhyans-MacBook-Pro / % hadoop fs -get /BDA/Sem_7_Timetable.csv /Users
2023-08-19 15:40:08,902 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
get: /Users/Sem_7_Timetable.csv._COPYING_ (Permission denied)
(base) dhyan@Dhyans-MacBook-Pro / % hadoop fs -cat /BDA/Sem_7_Timetable.csv
2023-08-19 15:40:38,643 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
Day,8:45-9:40,9:45-10:40,11:00-11:55,12:00-12:55,2:00-2:55,3:00-3:55,4:15-5:10,5:15-6:10
MON,,,"2CLOE02 RS & GIS
N308",,"EL1-IRS .
JP N305","EL2-CC .
SB N308","UEIM007 Div C
E103",
TUE,,,"2CLOE02 RS & GIS
N306",,"THOC
SB N307",,"UEIM007 Div C
E103",
WED,EL3-LAMP DV N507 EL4,,"THOC
SB N307",,"EL3-LAMP .
DV N305","EL2-CC .
SB N308",PROGCODE SCM N503 C2,
THU,,,"EL1-IRS .
JP N305","EL2-CC .
SB N308","THOC CT2
AV N306",,,
FRI,EL1-IRS JP N5111 EL3,,EL2-CC SB N5062 EL5,,"THOC
SB N307","EL3-LAMP .
DV N305",,
SAT,,,,,,,,
,,,,,,,,
,,,,,,,,
,,,,,,,,
```