Ex No – 9	Setting Up Hadoop Single Node Cluster

Aim:

To setup hadoop single node cluster

PROCEDURE:

Creating a Hadoop user for accessing HDFS and MapReduce

1. Create a new Hadoop user group and user account to deal with all Hadoop related activities.
2. Also, add the newly created Hadoop user account 'hduser' to the sudoers list to allow it to issue commands using sudo.

```
cloudlab@cloudlab...9020:~$sudo addgroup hadoop
cloudlab@cloudlab...9020:~$sudo adduser --ingroup hadoop
hduser cloudlab@cloudlab...9020:~$sudo adduser hduser sudo
```

Installing SSH

1. Install SSH to allow hadoop to access other nodes to start and manage all HDFS and Map Reduce Daemons.

```
cloudlab@cloudlab...9020:~$sudo apt-get install openssh-server
```

Configuring SSH

1. Login with 'hduser'
2. Generate ssh key for the user account
3. Copy the generated ssh key to the authorized keys from hduser

```
cloudlab@cloudlab...9020:~$sudo su hduser
hduser@cloudlab...9020:~$ssh-keygen -t rsa -P ""
hduser@cloudlab...9020:~$cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

Disabling Ipv6

1. Update the file /etc/sysctl.conf by adding the following line of codes at end of the file,

```
# disable ipv6
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv6.conf.default.disable_ipv6 = 1
net.ipv6.conf.lo.disable_ipv6 = 1
```

Installing Java

1. Download "jdk-7u79-linux-x64.tar.gz" to the Downloads directory.

2. Extract the downloaded file.
3. The contents are extracted in a new directory “jdk1.7.0_79”.
4. Move the directory to the location “/usr/local/jdk1.7.0_79”

```
hduser@cloudlab...9020:~$cd Downloads
hduser@cloudlab...9020:~Downloads$tar -xvzf jdk-7u79-linux-x64.tar.gz
hduser@cloudlab...9020:~Downloads$cd jdk1.7.0_79
hduser@cloudlab...9020:~Downloads/jdk1.7.0_79$sudo mv * /usr/local/jdk1.7.0_79
```

5. Update the bashrc file of ‘hduser’ account with the following java environment variables.

```
export JAVA_HOME=/usr/local/jdk1.7.0_79
export PATH=$PATH:$JAVA_HOME/bin
```

6. Now, source the bashrc file.

```
hduser@cloudlab...9020:~$source $HOME/.bashrc
```

7. Check the Java version Installed.

```
hduser@cloudlab...9020:~$java -version
java version "1.7.0_79"
Java(TM) SE Runtime Environment (build 1.7.0_79-b15)
Java HotSpot(TM) 64-Bit Server VM (build 24.79-b02, mixed mode)
```

Installing Hadoop

1. Download “hadoop-2.6.4.tar.gz” to the Downloads directory.
2. Extract the downloaded file.
3. The contents are extracted in a new directory “hadoop-2.6.4”
4. Move the contents of the directory “hadoop-2.6.4” to the location “/usr/local/hadoop”
5. Assign the Ownership of the directory “/usr/local/hadoop” to ‘hduser’
6. Create Hadoop temp directories for Namenode and Datanode
7. Again, assign Ownership of Hadoop Temp directories to ‘hduser’

```
hduser@cloudlab...9020:~$cd Downloads
hduser@cloudlab...9020:~Downloads$tar -xvzf  hadoop-2.6.4.tar.gz
hduser@cloudlab...9020:~Downloads$cd hadoop-2.6.4
hduser@cloudlab...9020:~Downloads/hadoop-2.6.4$sudo mv * /usr/local/hadoop
hduser@cloudlab...9020:~$sudo chown hduser:hadoop -R /usr/local/hadoop/
hduser@cloudlab...9020:~$sudo mkdir -p /usr/local/hadoop_tmp/hdfs/namenode
hduser@cloudlab...9020:~$sudo mkdir -p /usr/local/hadoop_tmp/hdfs/datanode
```

```
hduser@cloudlab...9020:~$sudo chown hduser:hadoop -R /usr/local/hadoop_tmp/
```

8. Update the bashrc file of 'hduser' account with the following Hadoop Environmental variables.

```
# -- HADOOP ENVIRONMENT VARIABLES START -- #
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"
export HADOOP_CLASSPATH= $JAVA_HOME/lib/tools.jar
# -- HADOOP ENVIRONMENT VARIABLES END -- #
```

9. Now, source the bashrc file.

```
hduser@cloudlab...9020:~$source $HOME/.bashrc
```

10. Check the Hadoop version Installed.

```
hduser@cloudlab...9020:~$hadoop version
Hadoop 2.6.4
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -
r 5082c73637530b0b7e115f9625ed7fac69f937e6 Compiled by
jenkins on 2016-02-12T09:45Z
Compiled with protoc 2.5.0
From source with checksum 8dee2286ecdbbbc930a6c87b65cbc010
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-
common-2.6.4.jar
```

Configuring hadoop-env.sh

1. Execute the below command to edit the file.

```
hduser@cloudlab...9020:~$ sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

2. Update the file with the following Java Home variable, save and exit.

```
## Update JAVA_HOME variable,
export JAVA_HOME=/usr/local/jdk1.7.0_79
```

Configuring core-site.xml

1. Execute the below command to edit the file.

```
hduser@cloudlab...9020:~$ sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml
```

2. Update the file with the following between the configuration tags, save and exit.

```
## Paste these lines into <configuration> tag
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>
```

Configuring hdfs-site.xml

1. Execute the below command to edit the file.

```
hduser@cloudlab...9020:~$ sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

2. Update the file with the following between the configuration tags, save and exit.

```
## Paste these lines into <configuration> tag
<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>
```

Configuring yarn-site.xml

1. Execute the below command to edit the file.

```
hduser@cloudlab...9020:~$ sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml
```

2. Update the file with the following between the configuration tags, save and exit.

```
## Paste these lines into
<configuration> tag <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>dfs.datanode.data.dir</name>
  <value>file:/usr/local/hadoop_tmp/hdfs/datanode</value>
</property>
```

Configuring mapred-site.xml

1. Copy the template of mapred-site.xml.template to mapred-site.xml and edit the maprd-site.xml file as follows,

```
hduser@cloudlab...9020:~$ sudo cp /usr/local/hadoop/etc/hadoop/mapred-
site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml
hduser@cloudlab...9020:~$ sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml
```

2. Update the file with the following between the configuration tags, save and exit.

```
## Paste these lines into <configuration> tag
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

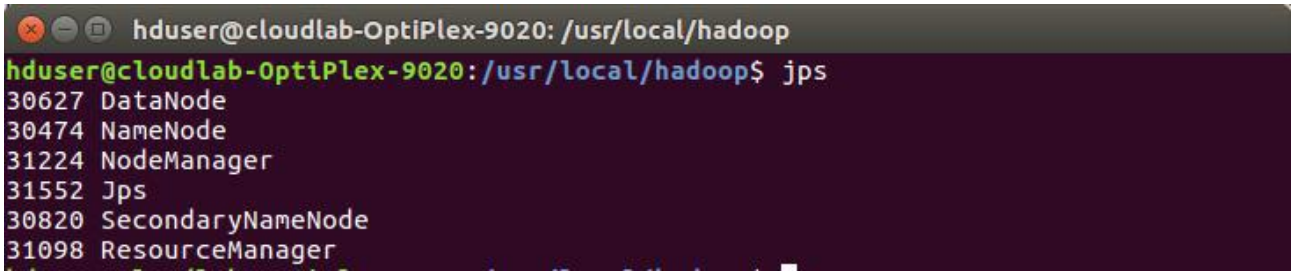
Format Namenode

```
hduser@cloudlab...9020:~$ hdfs namenode -format
```

Start all Hadoop Daemons

```
hduser@cloudlab...9020:~$ cd $HADOOP_HOME
hduser@cloudlab...9020:~/usr/local/hadoop$ start-dfs.sh
hduser@cloudlab...9020:~/usr/local/hadoop$ start-yarn.sh
```

Track/Monitor/Verify



Output:

Thus the Hadoop single node Cluster setup have been successfully done and all the hadoop daemons are verified to be running properly.