CDH4 Pseudo installation on CentOS

Note: **Do not create username as hadoop as you will have issues in installation.**

1) Install Java

Copy the Oracle jdk on to desktop.

```
$ cd Desktop
$ su
$ rpm -i jdk-7u67-linux-i586.rpm
```

2) Install CDH4 repository

Install CDH4 repository

Click on links : for 32 bit link

```
cd /home/username/Downloads/
rpm -i cloudera-*.rpm
```

3) Install CDH4 with MRv2

```
yum install hadoop-0.20-mapreduce-jobtracker -y
yum install hadoop-hdfs-namenode -y
yum install hadoop-hdfs-secondarynamenode -y
yum install hadoop-0.20-mapreduce-tasktracker  hadoop-hdfs-datanode -y
```

4) Set Java and Hadoop Home (Environmental variables for Linux)

Using command:

```
gedit  /etc/profile
# Set Hadoop_Home
export HADOOP_HOME=/usr/lib/hadoop
export PATH=$PATH:/usr/lib/hadoop/bin
```

To test the above settings:

```
whereis hadoop
```

Configuration (Note: All following steps up to 'User Assignment' must be done only in /etc/hadoop/conf directory)

Go to hadoop config directory

cd /etc/hadoop/conf

5) Add the below property tags in between <configuration> tags

gedit core-site.xml

```
<property>
<name>hadoop.tmp.dir</name>
<value>/usr/lib/hadoop/tmp</value>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:8020</value>
</property>
```

```
mkdir /usr/lib/hadoop/tmp
chmod 777 /usr/lib/hadoop/tmp/
chown hdfs:hadoop /usr/lib/hadoop/tmp/
```

6) gedit hdfs-site.xml

```
<property>
<name>dfs.permissions</name>
<value>false</value>
</property>
<property>
<name>dfs.name.dir</name>
<value>/storage/name</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/storage/data</value>
</property>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
```

```
mkdir /storage
chmod 775 /storage/
chown hdfs:hadoop /storage/
```

7) gedit mapred-site.xml

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>hdfs://localhost:8021</value>
</property>
<property>
<name>mapred.system.dir</name>
<value>/mapred/system</value>
</property>
<property>
<name>mapred.local.dir</name>
<value>/mapred/local</value>
</property>
<property>
<name>mapred.temp.dir</name>
<value>/mapred/temp</value>
</property>
</configuration>
```

```
mkdir /mapred
chmod 775 /mapred
chown mapred:hadoop /mapred
```

8) User Assignment

```
export HADOOP_NAMENODE_USER=hdfs
export HADOOP_SECONDARYNAMENODE_USER=hdfs
export HADOOP_DATANODE_USER=hdfs
export HADOOP_JOBTRACKER_USER=mapred
export HADOOP_TASKTRACKER_USER=mapred
```

9) Format namenode

```
sudo -u hdfs hdfs namenode -format
```

You must get a successfully formatted message. Otherwise, check the /var/log/hadoop-hdfs/ directories for error logs and correct them.

10) Start Daemons

/etc/init.d/hadoop-hdfs-namenode start
/etc/init.d/hadoop-hdfs-secondarynamenode start
/etc/init.d/hadoop-0.20-mapreduce-jobtracker start
/etc/init.d/hadoop-hdfs-datanode start
/etc/init.d/hadoop-0.20-mapreduce-tasktracker start

If any daemon fails to start. Check for any errors in /var/log/hadoop-hdfs/ and /var/log/hadoop-0.20-mapreduce/ directories for respective daemon

11) Check WebUI

localhost:50070 – Namenode (HDFS)
localhost:50030 – Jobtracker (MapReduce)
localhost:50090 – Seconadary Namenode