Hadoop Cluster Setup on Linux (CentOS 6.5) Hosts.

1. Create hadoop specific linux group and user and set password

```
$groupadd hduser
$adduser -g hduser hduser
$passwd hduser
```

2. Install JDK (Oracle Jdk 1.8) on all machines

3. Maping the nodes

3.1 edit hosts file in /etc/ folder on all nodes, specify the IP address of each system followed by their host names.

```
$ vi /etc/hosts

192.168.1.100 hadoop-master
192.168.1.101 hadoop-slave-1
192.168.1.102 hadoop-slave-2
192.168.1.103 hadoop-slave-3
```

3.2 If required, change all machines' host names by editing /etc/sysconfig/network file on each machine and set the host name as mentioned in **/etc/hosts**.

4. Configuring key based login

Setup ssh in every node such that they can communicate with one another without any prompt for password. Execute following commands on each node.

```
$ ssh-keygen -t rsa  -P
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@hadoop-master
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@hadoop-slave-1
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@hadoop-slave-2
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@hadoop-slave-3
$ chmod 0600 ~/.ssh/authorized_keys
```

5. Installing Hadoop

On the master node, download and install Hadoop using the following commands. (Run this commands as root).

```
$wget http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.8.0/hadoop-2.8.0.tar.gz
$tar -xvf hadoop-2.8.0.tar.gz
$mv hadoop-2.8.0 /opt/hadoop
$chown -R hduser:hduser hadoop
```

6. Configuring Hadoop

6.1 Edit the following files given insdie **/opt/hadoop/etc/hadoop**.

core-site.xml

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop-master:9000/</value>
  </property>
```

```
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/opt/hadoop/tmp</value>
  </property>
</configuration>
```

hdfs-site.xml

On hadoop-master machine

```
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/opt/hadoop/tmp/dfs/name</value>
  </property>

  <property>
    <name>dfs.hosts</name>
    <value>/opt/hadoop/etc/hadoop/datanodes.lst</value>
  </property>

  <property>
    <name>dfs.blocksize</name>
    <value>33554432</value>
  </property>

  <property>
    <name>dfs.namenode.handler.count</name>
    <value>10</value>
  </property>

  <property>
    <name>dfs.namenode.http-address</name>
    <value>hadoop-master:50070</value>
  </property>

  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>hadoop-master:50090</value>
  </property>
</configuration>
```

On hadoop-slave machines (do this after copying hadoop directory on slave nodes i.e. at the end of step 8)

```
<configuration>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/opt/hadoop/tmp/dfs/data</value>
  </property>
</configuration>
```

mapred-site.xml

Create mapred-site.xml by copying it from mapred-site.xml.template

```
$cp mapred-site.xml.template mapred-site.xml
```

Edit the file mapred-site.xml

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

  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>hadoop-master:10020</value>
```

```
  </property>

  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>hadoop-master:19888</value>
  </property>
</configuration>
```

yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>hadoop-master</value>
  </property>

  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

6.2 create a file named **"slaves"** (only on master node) with following lines as content

```
hadoop-slave-1
hadoop-slave-2
hadoop-slave-3
```

6.3 create a file named **"datanodes.lst"** (only on master node) with following lines as content

```
hadoop-slave-1
hadoop-slave-2
hadoop-slave-3
```

7. Set environment variables

7.1 Edit files **hadoop-env.sh**, **mapred-env.sh** and **yarn-env.sh** inside **/opt/hadoop/etc/hadoop** directory, with following line to set environment variable JAVA_HOME

```
export JAVA_HOME=/usr/java/jdk1.8.0_91
```

7.2 Change **/home/hduser/.bashrc** by appending the following line in it.

```
export JAVA_HOME=/usr/java/jdk1.8.0_91
export HADOOP_HOME=/opt/hadoop
export HADOOP_PREFIX=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

7.3 run following command to bring all the environment variables in effect.

```
$source ~/.bashrc
```

8. Install hadoop on slave machines

Run following commands from the **hadoop-master** node

```
$cd /opt
$scp -r hadoop hadoop-slave-1:/opt/
$scp -r hadoop hadoop-slave-2:/opt/
$scp -r hadoop hadoop-slave-3:/opt/
```

```
$scp /home/hduser/.bashrc hadoop-slave-1:/home/hduser
$scp /home/hduser/.bashrc hadoop-slave-2:/home/hduser
$scp /home/hduser/.bashrc hadoop-slave-3:/home/hduser

$ssh hduser@hadoop-slave-1
$source ~/.bashrc

$ssh hduser@hadoop-slave-2
$source ~/.bashrc

$ssh hduser@hadoop-slave-3
$source ~/.bashrc
```

9. Start the hadoop cluster from hadoop-master node

Format the **namenode (only once)**

```
$ mkdir -p /opt/hadoop/tmp/dfs/name
$ mkdir -p /opt/hadoop/tmp/dfs/data


$hdfs namenode -format
```

Start the **namenode**, **secondarynamenode** and **datanodes**

```
$sh start-dfs.sh
```

Start the **resourcemanager** and **nodemanagers**

```
$sh start-yarn.sh
```

Start the **MapReduce JobHistory server**

```
$sbin/mr-jobhistory-daemon.sh --config $HADOOP_CONF_DIR start historyserver
```

10. Stop the hadoop cluster from hadoop-master node

Stop the **namenode**, **secondarynamenode** and **datanodes**

```
$sh stop-dfs.sh
```

Stop the **resourcemanager** and **nodemanagers**

```
$sh stop-yarn.sh
```

Stop the **MapReduce JobHistory server**

```
$sbin/mr-jobhistory-daemon.sh --config $HADOOP_CONF_DIR stop historyserver
```
