Hadoop multi-nodes Installation

Environment:

Hadoop 2.7.2

Ubuntu 14.04 LTS

ssh-keygen

Java version 1.8.0

Scala 2.11.7

Servers:

Master: 192.168.199.80 (hadoopmaster)

Hadoopslave: 192.168.199.81(hadoopslave1)

Hadoopslave: 192.168.199.82(hadoopslve2)

Install Java 8:

```
sudo add-apt-repository ppa:openjdk-r/ppa
sudo apt-get update
sudo apt-get install openjdk-8-jdk
sudo update-alternatives --config java
sudo update-alternatives --config javac
```

Add JAVA_HOME to ~/.bashrc

```
$ sudo vi ~/.bashrc
//add two lines at the end of .bashrc
export JAVA_HOME=/usr/lib/java-8-openjdk-amd64
export PATH=PATH:$JAVA_HOME/bin
```

Then source it

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

Tips:

Don't forget it is a hidden file inside your home directory (you would not be the first to do a ls -l and thinking it is not there).

```
ls -la ~/ | more
```

ADD Hosts

```
# vi /etc/hosts

enter the following lines in the /etc/hosts file.

192.168.199.80 hadoopmaster

192.168.199.81 hadoopslave1

192.168.199.82 hadoopslave2
```

Setup SSH in every node

So they can communicate without password (do the same in three nodes)

```
$ ssh-keygen -t rsa

$ ssh-copy-id -i ~/.ssh/id_rsa.pub cmtadmin@hadoopmaster

$ ssh-copy-id -i ~/.ssh/id_rsa.pub cmtadmin@hadoopslave1

$ ssh-copy-id -i ~/.ssh/id_rsa.pub cmtadmin@hadoopslave2

$ chmod 0600 ~/.ssh/authorized_keys

$ exit
```

Install Hadoop 2.7.2 (to /opt/Hadoop)

Download from Hadoop 2.7.2(Hadoop-2.7.2.tar.gz)

Hadoop-2.7.2-src.tar.gz is the version you need to build by yourself

```
$ tar xvf Hadoop-2.7.2.tar.gz  /opt

$ cd /opt/hadoop
```

Configuring Hadoop

core-site.xml

Open the core-site.xml file and edit it as shown below.

```
<configuration>

   <property>

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

      <value>hdfs://hadoopmaster:9000/</value>

   </property>

   <property>

      <name>dfs.permissions</name>
```

```
      <value>false</value>
   </property>
</configuration>
```

hdfs-site.xml

Open the hdfs-site.xml file and edit it as shown below.

```
<configuration>
   <property>
      <name>dfs.data.dir</name>
      <value>/media/hdfs/name/data</value>
      <final>true</final>
   </property>
   <property>
      <name>dfs.name.dir</name>
      <value>/media/hdfs/name</value>
      <final>true</final>
   </property>
   <property>
      <name>dfs.replication</name>
      <value>1</value>
   </property>
</configuration>
```

mapred-site.xml

Open the mapred-site.xml file and edit it as shown below.

```
<configuration>
   <property>
      <name>mapred.job.tracker</name>
      <value>hadoopmaster:9001</value>
   </property>
</configuration>
```

hadoop-env.sh

Open the hadoop-env.sh file and edit JAVA_HOME

Installing Hadoop on Slave Servers

```
$ cd /opt
```

```
$ scp -r hadoop hadoopslave1:/opt/
$ scp -r hadoop hadoopslave2:/opt/
```

Configuring Hadoop on Master Server

```
$ cd /opt/hadoop
$ vi etc/hadoop/masters
hadoopmaster
$ vi etc/hadoop/slaves
hadoopslave1
hadoopslave2
```

Add HADOOP_HOME, PATH

```
export HADOOP_HOME=/opt/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
```

Format Name Node on Hadoop Master

```
$ cd /opt/hadoop/hadoop
$ bin/hadoop namenode –format
```

Start Hadoop services

```
$ cd /opt/hadoop/sbin
$ start-all.sh
```

Stop all the services

```
$ cd /opt/hadoop/sbin
$ stop-all.sh
```

Installation Spark 1.6 based on user-provided Hadoop

Step 1 install scala

Install Scala 2.11.7 download from website

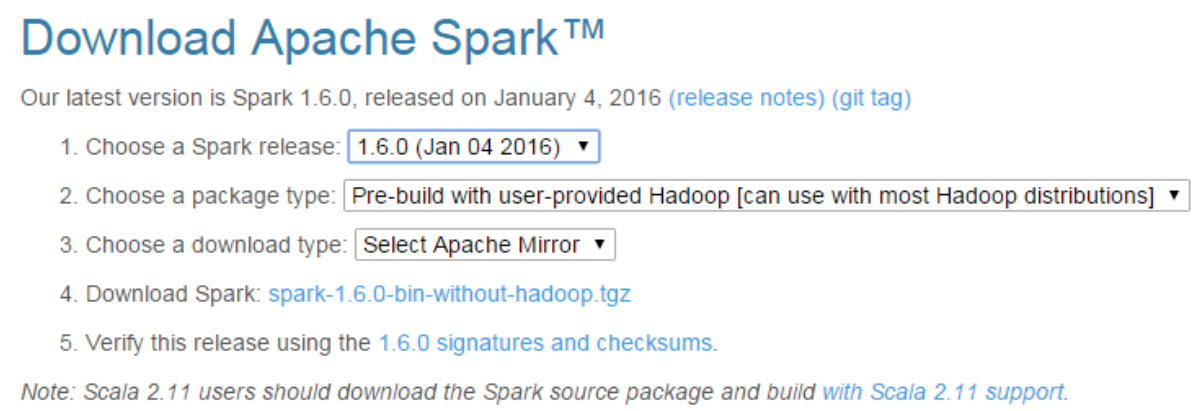
```
$ tar xvf scala-2.11.7.tgz
$ mv scala-2.11.7/ /usr/opt/scala
```

Set PATH for Scala in ~/.bashrc

```
$ sudo vi ~/.bashrc
```

```
export SCALA_HOME=/usr/opt/scala
export PATH = $PATH:$SCALA_HOME/bin
```

Download Spark 1.6 from apache server



Install Spark

```
$ tar xvf spark-1.6.0-bin-without-hadoop.tgz
$ mv spark-1.6.0-bin-without-hadoop/  /opt/spark
```

Set up environment for spark

```
$ sudo vi ~/.bashrc
 export SPARK_HOME=/usr/opt/spark
 export PATH = $PATH:$SPARK_HOME/bin
```

Add entity to configuration

```
$ cd /opt/spark/conf
$ cp spark_env.sh.template spark_env.sh
$ vi spark_env.sh
HADOOP_CONF_DIR=/opt/hadoop/etc/hadoop
export SPARK_DIST_CLASSPATH=$(hadoop classpath)
```

Add slaves to configuration

```
$ cd /opt/spark/conf
$ cp slaves.template slaves
$ vi slaves
hadoopslave1
```

```
hadoopslave2
```

Run spark

```
$ cd /opt/spark/bin
$ spark-shell
```