Installation of hadoop 2.2.x

hjw

September 14, 2014

1 Prerequiste

1.1 Install Ubuntu

Via wubi, virtual box, vmware station, etc. Install it the way you like.

1.2 Install Java and ssh-server

```
$ sudo apt-get install openjdk-7-jdk
$ java -version
```

It should be something like:
java version "1.7.0_65"
OpenJDK Runtime Environment ...

```
$ cd /usr/lib/jvm
$ sudo ln -s java-7-openjdk-i386 jdk
# sudo apt-get install openssh-server
```

2 Create Hadoop User

```
$ sudo addgroup hadoop
$ sudo adduser --ingroup hadoop hduser
$ sudo adduser hduser sudo
```

3 Setup SSH Certification

Now, re-login into ubuntu using hduser.

```
$ ssh-keygen -t rsa -P ''
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ ssh localhost
```

4 Download Hadoop 2.2.0

```
$ cd ~
$ wget http://www.trieuvan.com/apache/hadoop/common/hadoop-2.2.0/hadoop-2.2.0.tar.gz
$ sudo tar vxzf hadoop-2.2.0.tar.gz -C /usr/local
$ cd /usr/local
$ sudo mv hadoop-2.2.0 hadoop
$ sudo chown -R hduser:hadoop hadoop
```

5 Setup Hadoop Enironment

```
$cd ~
$vi .bashrc

paste following to the end of the file

#Hadoop variables
export JAVA_HOME=/usr/lib/jvm/jdk/
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
###end of paste

$ cd /usr/local/hadoop/etc/hadoop
$ vi hadoop-env.sh

#modify JAVA_HOME
export JAVA_HOME=/usr/lib/jvm/jdk/
```

6 Configure Hadoop

Now, re-login into ubuntu using hduser

```
$ hadoop version
Hadoop 2.2.0
Subversion https://svn.apache.org/repos/asf/hadoop/common -r 1529768
Compiled by hortonmu on 2013-10-07T06:28Z
Compiled with protoc 2.5.0
From source with checksum 79e53ce7994d1628b240f09af91e1af4
This command was run using /usr/local/hadoop-2.2.0/share/hadoop/common/hadoop-common-2.2.0.jar
```

If you get similar printing, then hadoop is installed.

```
$ cd /usr/local/hadoop/etc/hadoop
$ vi core-site.xml
#Paste following between <configuration>

<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
</property>


$ vi yarn-site.xml
#Paste following between <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>


$ mv mapred-site.xml.template mapred-site.xml
$ vi mapred-site.xml
#Paste following between <configuration>

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

```
$ cd ~
$ mkdir -p mydata/hdfs/namenode
$ mkdir -p mydata/hdfs/datanode
$ cd /usr/local/hadoop/etc/hadoop
$ vi hdfs-site.xml
Paste following between <configuration> tag

<property>
   <name>dfs.replication</name>
   <value>1</value>
 </property>
 <property>
   <name>dfs.namenode.name.dir</name>
   <value>file:/home/hduser/mydata/hdfs/namenode</value>
 </property>
 <property>
   <name>dfs.datanode.data.dir</name>
   <value>file:/home/hduser/mydata/hdfs/datanode</value>
 </property>
```

7 Format Namenode

```
$ hdfs namenode -format
```

8 Start Hadoop Service

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

$ jps

If everything is sucessful, you should see following services running
2583 DataNode
2970 ResourceManager
3461 Jps
3177 NodeManager
2361 NameNode
2840 SecondaryNameNode
```

9 Stop Services

You can stop hadoop services and exit sshserver as follows:

```
$ stop-yarn.sh
$ stop-dfs.sh
$ exit
```