

# Installation of hadoop 2.2.x

---

## 1 Prerequisite

---

### 1.1 Install Ubuntu

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

注：进行本次实验前，建议将分配给虚拟机的cpu数和ram尽量大（virtualbox中在 控制-设置-系统 中设置）。

若ram分配不足，后续测试map reduce程序时可能会运行得非常慢。

### 1.2 Create Hadoop User

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

### 1.3 Setup SSH Certification

**Now, re-login into ubuntu using hduser.**

```
ssh hduser@localhost
```

输入密码后，即使用hduser账户登陆成功后，配置免密登陆：

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

使用以下命令再进行一次免密登陆：

```
$ ssh localhost
```

**注意，之后的命令，都应在hduser账户下执行，命令行提示符前应该显示hduser@hostname。**（注：上述配置完成后，仅从hduser账户从本机登陆hduser是免密的，其他情况仍要密码登陆的）

### 1.4 Install Java and ssh-server

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

It should be something like:

```
java version "1.7.0_151"
OpenJDK Runtime Environment ...
```

Then

```
$ cd /usr/lib/jvm
$ sudo ln -s java-7-openjdk-amd64 jdk
$ sudo apt-get install openssh-server
```

(注：在32位系统内，上述第二行命令中的amd64可能需要修改为i386)

## 2 Download Hadoop 2.2.0

---

```
$ cd ~
$ wget http://archive.apache.org/dist/hadoop/core/hadoop-2.2.0/hadoop-2.2.0.tar.gz
```

可以参考<https://teddysun.com/377.html>，下载并安装一个axel提高下载速度，安装好后，输入命令：

```
axel http://archive.apache.org/dist/hadoop/core/hadoop-2.2.0/hadoop-2.2.0.tar.gz -n 32
```

下载完成后，执行：

```
$ 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
```

## 3 Setup Hadoop Environment

---

```
$ 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
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
# end of paste
```

Then

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

modify JAVA\_HOME

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

如果显式只读无法修改的话，在上一步输入：

```
$ sudo vi hadoop-env.sh
```

## 4 Configure Hadoop

**Now, re-login into ubuntu using hduser**

这一步是为了使上面设置的.bashrc生效； 由于之前通过ssh localhost登陆过，此时可依次输入

```
$ exit
$ ssh localhost
```

来重新登陆。

```
$ 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> and </configuration> tag
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>
```

```
$ vi yarn-site.xml
```

```
# Paste following between <configuration> and </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>
```

```
$ mv mapred-site.xml.template mapred-site.xml
$ vi mapred-site.xml
```

```
# Paste following between <configuration> and </configuration> tag
<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> and </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>
```

## 5 Format Namenode

---

```
$ hdfs namenode -format
```

(注：这一步不要重复执行)

## 6 Start Hadoop Service

---

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

(注：中途可能需要输入yes)

```
$ jps
```

If everything is successful, you should see following services running

```
2583 DataNode
2970 ResourceManager
3461 Jps
3177 NodeManager
2361 NameNode
2840 SecondaryNameNode
```

注：服务名前的数字无所谓。

注：若DataNode未启动，可能是由于重复执行了hdfs namenode -format，此时应先删除并重建hdfs，即依次执行：

```
$ stop-dfs.sh
$ rm -rf /home/hduser/mydata
$ mkdir -p /home/hduser/mydata/hdfs/namenode
$ mkdir -p /home/hduser/mydata/hdfs/datanode
$ hdfs namenode -format
$ start-dfs.sh
```

## 7 Stop Services

---

You can stop hadoop services and exit sshserver as follows: 若要停止hadoop相关服务：

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

可再用jps命令查看相关服务是否停止：

```
$ jps
3461 Jps
```