

Ubuntu Setup

- Running Linux on your Mac: 2013 edition ([link](#))

- Download Ubuntu LTS version (long-term support) <http://www.ubuntu.com/download/desktop>
- Install Guest-Addition <https://help.ubuntu.com/community/VirtualBox/GuestAdditions>
 - VBoxGuestAdditions_4.3.20.iso

Hadoop 2 Installation on Ubuntu

- Hadoop V1 Installation on Ubuntu ([link](#))

OpenSSH

```
$ sudo apt-get install openssh-server
```

Update software

```
$ sudo apt-get update
```

Java

Download oracle java JDK to ~/Downloads

```
$ cd /usr/local
$ sudo mkdir java
$ cd java
$ cd ~/Downloads
$ sudo tar xvzf ~/Downloads/jdk-7u75-linux-x64.tar.gz
$ sudo mv jdk_1.7.0_75 /usr/local/java/
```

Edit system PATH file

```
$ sudo gedit ~/.bash_profile

export JAVA_HOME=/usr/local/Java/jdk1.7.0_75
export PATH=$PATH:$HOME/bin:$JAVA_HOME/bin

$. ~/.bash_profile

$ java -version
```

SSH

```
$ ssh-keygen -t dsa -P "" -f ~/.ssh/id_dsa
$ cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys

$ ssh hadoop@localhost
```

Disable IPV6

```
$ sudo gedit /etc/sysctl.conf

-- add line
#disable ipv6
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv6.conf.default.disable_ipv6 = 1
net.ipv6.conf.lo.disable_ipv6 = 1
```

Hadoop

- \$ curl -O <http://apache.mirror.rafael.ca/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz>
- \$ tar -zxvf hadoop-2.6.0.tar.gz
- \$ sudo mv ~/Downloads/hadoop-2.6.0 /usr/local/hadoop-2.6.0

Configuration

```
$ cd /usr/local/hadoop-2.6.0/etc/hadoop
$ll
```

a. yarn-site.xml:

```
<configuration>
<!-- Site specific YARN configuration properties -->
<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>
</configuration>
```

b. core-site.xml

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

c. mapred-site.xml

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

d. hdfs-site.xml

```
$ sudo mkdir -p $HADOOP_HOME/yarn_data/hdfs/namenode
$ sudo mkdir -p $HADOOP_HOME/yarn_data/hdfs/datanode
$ sudo chown hadoop /usr/local/hadoop-2.6.0
$ sudo chgrp hadoop /usr/local/hadoop-2.6.0
```

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>file:/usr/local/hadoop-2.6.0/yarn_data/hdfs/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>file:/usr/local/hadoop-2.6.0/yarn_data/hdfs/datanode</value>
</property>
</configuration>
```

e. Update \$HOME/.bashrc

```
# Set Hadoop-related environment variables
export HADOOP_PREFIX=/usr/local/hadoop
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_MAPRED_HOME=${HADOOP_HOME}
export HADOOP_COMMON_HOME=${HADOOP_HOME}
export HADOOP_HDFS_HOME=${HADOOP_HOME}
export YARN_HOME=${HADOOP_HOME}
export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop

# Native Path
export HADOOP_COMMON_LIB_NATIVE_DIR=${HADOOP_PREFIX}/lib/native
export HADOOP_OPTS="-Djava.library.path=${HADOOP_PREFIX}/lib"

#Java path
export JAVA_HOME='/usr/local/Java/jdk1.7.0_75'

# Add Hadoop bin/ directory to PATH
export PATH=$PATH:$HADOOP_HOME/bin:$JAVA_PATH/bin:$HADOOP_HOME/sbin
```

Launch Hadoop

```
$ bash hadoop namenode -format
$ bash hadoop-daemon.sh start namenode
$ bash hadoop-daemon.sh start datanode
$ bash yarn-daemon.sh start resourcemanager
$ bash yarn-daemon.sh start nodemanager
$ bash mr-jobhistory-daemon.sh start historyserver
```

Access web interfaces

- Cluster status: <http://localhost:8088>
- HDFS status: <http://localhost:50070>
- Secondary NameNode status: <http://localhost:50090>