Setting up a Single Node Hadoop Cluster on Ubuntu 14.04

Patrick Loftus

This guide documents the steps I took to set up an apache hadoop single node cluster on Ubuntu 14.04. Most of these steps are taken from the following online resources:

- [http://www.bogotobogo.com/Hadoop/BigData_hadoop_Install_on_ubuntu_single_node_cluster.php](http://www.bogotobogo.com/Hadoop/BigData_hadoop_Install_on_ubuntu_single_node_cluster.php)

**Install Java**

- `sudo apt-get update`
- `sudo apt-get install default-jdk`

**Add a dedicated hadoop user and add to groups**

- `sudo add group hadoop`
- `sudo adduser -ingroup hadoop hduser`
- `sudo adduser hduser sudo`

**Install and configure SSH**

- `sudo apt-get install ssh`
- `su hduser (you’re now logged in as the hduser)`
- `ssh-keygen -t rsa -P””`

**Install hadoop**

- `wget http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz`
- `tar xvzf hadoop-2.6.0.tar.gz`
- `sudo mkdir /usr/local/hadoop`
- `sudo mv hadoop-2.6.0/* /usr/local/hadoop`
- `sudo rm -r hadoop-2.6.0`
- `sudo chown -R hduser:hadoop /usr/local/hadoop`

**Setup configuration files**

- add the following to the end of your .bashrc file
  - `export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64`
  - `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_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native`
• export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"

• source .bashrc
• sudo mkdir -p /app/hadoop/tmp
• sudo chown hduser:hadoop /app/hadoop/tmp
• specify the following configuration in /usr/local/hadoop/etc/hadoop/core-site.xml:

```xml
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/app/hadoop/tmp</value>
    <description>A base for other temporary directories.</description>
  </property>

  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:54310</value>
    <description>The name of the default file system. A URI whose scheme and authority determine the FileSystem implementation. The uri's scheme determines the config property (fs.SCHEME.impl) naming the FileSystem implementation class. The uri's authority is used to determine the host, port, etc. for a filesystem.</description>
  </property>
</configuration>
```

• cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml
• add the following to /usr/local/hadoop/etc/hadoop/mapred-site.xml

```xml
<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>localhost:54311</value>
    <description>The host and port that the MapReduce job tracker runs at. If "local", then jobs are run in-process as a single map and reduce task.</description>
  </property>
</configuration>
```

• tell hadoop where your java files are by adding the following line to /usr/local/hadoop/conf/hadoop-env.sh:

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
```

Start Hadoop

• format the name node:

```
/usr/local/hadoop$ bin/hadoop namenode -format
```

• if everything went as planned you should be able to start hadoop with:
start-all.sh
• the jps command will give you the ids of virtual nodes and the names of the daemons they are running
• stop-all.sh will shut down the virtual hadoop cluster