Hadoop – 2.6.0 Installation

Downloading from the website and installing the Hadoop executable:

(1) Visit http://hadoop.apache.org; go to "Getting Started"; and click "Download".
(2) From the "Hadoop release" page, click "Download".
(3) From the "Download" page, click "Download release now".
(4) From the "The Apache Software Foundation" page, click any mirror site.
(5) From the "Hadoop Releases", click Hadoop-2.6.0.
(6) From the index of /hadoop/common/Hadoop-2.6.0, click "hadoop_2.6.0.tar.gz"
(7) Move "hadoop_2.6.0.tar.gz" to your home directory.
(8) gzip –d hadoop_2.6.0.tar.gz
(9) tar –xvf - < hadoop_2.6.0.tar

Or just copying from ~css534 and installing the Hadoop executable:

(1) cp ~css534/hadoop_2.6.0/hadoop.2.6.0.tar.gz ~/
(2) gzip –d hadoop_2.6.0.tar.gz
(3) tar –xvf - < hadoop_2.6.0.tar

Downloading the Hadoop user manual:

(1) Visit http://hadoop.apache.org/docs/ and click "r.2.6.0".
(2) From "Apache Hadoop 2.6.0" page, click "Cluster Setup".

Setting up configuration files under ~/hadoop_2.6.0/

(1) Configuring ~/hadoop_2.6.0/etc/hadoop/hadoop-env.sh

```
export JAVA_HOME=/usr/java/latest
export HADOOP_PREFIX=$HOME/hadoop-2.6.0
```

Or, you can simply copy ~css534's file to your directory:
cp ~css534/hadoop_2.6.0/etc/hadoop/hadoop-env.sh ~/hadoop_2.6.0/etc/hadoop/hadoop-env.sh

(2) Configuring the Hadoop Daemons in Non-Secure Mode
   a. ~/hadoop_2.6.0/etc/hadoop/core-site.xml

```
<configuration>
     <property>
          <name>fs.defaultFS</name>
          <value>hdfs://uw1-320-10:50763</value>
     </property>
</configuration>
```

Please change 50763 into the last 5 digits of your student ID (that should be 5001 through 65535). If not, choose your favorite number. Don't choose a simple number like 11111, 12345, 23232, etc.

Or, you can simply copy ~css534's file to your directory:
cp ~css534/hadoop_2.6.0/etc/hadoop/core-site.xml ~/hadoop_2.6.0/etc/hadoop/core-site.xml
However, please don't forget to change the port number 50763!

b. ~/hadoop_2.6.0/etc/hadoop/hdfs-site.xml

```
<configuration>
    <property>
       <name>dfs.replication</name>
       <value>1</value>
    </property>

   <property>
      <name>hadoop.tmp.dir</name>
      <value>/tmp/hadoop-css534</value>
      <final>true</final>
   </property>

   <property>
      <name>dfs.namenode.name.dir</name>
      <value>${hadoop.tmp.dir}/dfs/name</value>
      <final>true</final>
   </property>

   <property>
      <name>dfs.datanode.data.dir</name>
      <value>${hadoop.tmp.dir}/dfs/data</value>
      <final>true</final>
   </property>

   <property>
      <name>dfs.checkpoint.dir</name>
      <value>${hadoop.tmp.dir}/dfs/namesecondary</value>
      <final>true</final>
   </property>

   <property>
      <name>dfs.secondary.http.address</name>
      <value>0.0.0.0:50764</value>
   </property>

   <property>
      <name>dfs.datanode.address</name>
      <value>0.0.0.0:50765</value>
   </property>

   <property>
      <name>dfs.datanode.http.address</name>
      <value>0.0.0.0:50766</value>
   </property>

   <property>
      <name>dfs.datanode.ipc.address</name>
      <value>0.0.0.0:50767</value>
   </property>

   <property>
      <name>dfs.http.address</name>
      <value>0.0.0.0:50768</value>
   </property>
```

```
</configuration>
```

Please change /tmp/hadoop-css534 into /tmp/hadopp-YourUnetID.
For instance, if your UnetID is mfukuda, it should be /tmp/hadoop-mfukuda
Please change all port numbers:
50764: your original port + 1
50765: your original port + 2
50766: your original port + 3
50767: your original port + 4
50768: your original port + 5

Or, you can simply copy ~css534's file to your directory:
cp ~css534/hadoop_2.6.0/etc/hadoop/hdfs-site.xml ~/hadoop_2.6.0/etc/hadoop/hdfs-site.xml
However, please don't forget to change /tmp/hadoop-css534 as well as all port numbers such as 50764, 50765, 50766, 50767, and 50768!

For more details, please check:
http://hadoop.apache.org/docs/r2.4.0/hadoop-project-dist/hadoop-hdfs/hdfs-default.xml

c. ~/hadoop_2.6.0/etc/hadoop/mapred_site.xml

```
<configuration>

   <property>
      <name>hadoop.tmp.dir</name>
      <value>/tmp/hadoop-css534</value>
      <final>true</final>
   </property>

   <property>
      <name>mapred.job.tracker</name>
      <value>0.0.0.0:20369</value>
      <final>true</final>
   </property>

  <property>
      <name>mapred.local.dir</name>
      <value>${hadoop.tmp.dir}/mapred/local</value>
      <final>true</final>
   </property>

   <property>
      <name>mapred.system.dir</name>
      <value>${hadoop.tmp.dir}/mapred/system</value>
      <final>true</final>
   </property>

   <property>
      <name>mapred.tasktracker.map.tasks.maximum</name>
      <value>4</value>
      <final>true</final>
   </property>
```

```
    <property>
       <name>mapred.tasktracker.reduce.tasks.maximum</name>
       <value>4</value>
       <final>true</final>
    </property>

    <property>
       <name>mapred.child.java.opts</name>
       <value>-Xmx1g</value>
       <final>true</final>
    </property>

</configuration>
```

Please change /tmp/hadoop-css534 into /tmp/hadopp-YourUnetID.
For instance, if your UnetID is mfukuda, it should be /tmp/hadoop-mfukuda
Please change all port numbers:
    20369: your original port + 6

Or, you can simply copy ~css534's file to your directory:
cp ~css534/hadoop_2.6.0/etc/hadoop/mapred-site.xml ~/hadoop_2.6.0/etc/hadoop/map-site.xml
However, please don't forget to change /tmp/hadoop-css534 as well as the port number 20369!

d. ~/hadoop_2.6.0/etc/hadoop/slaves

```
uw1-320-10
uw1-320-11
uw1-320-12
uw1-320-13
```

Choose 4 computing nodes. The first node should be the same as the node that you specified in core-site.xml.
Or, you can simply copy ~css534's file to your directory:
cp ~css534/hadoop_2.6.0/etc/hadoop/slave ~/hadoop_2.6.0/etc/hadoop/slave

(3) Format a new file system (just one time)
Make sure that you logged in uw1-320-10 or the master node you decided.
cd ~/hadoop-2.6.0
./bin/hdfs namenode –format

(4) Start HDFS
./sbin/start-dfs.sh

(5) Create a new account
In the following example, you need to replace css534 with YourUNetID such as mfukuda
./bin/hdfs dfs –mkdir /user
./bin/hdfs dfs –mkdir /user/css534
./bin/hdfs dfs –chown css534:css534 /user/css534

./bin/hdfs dfs –mkdir /user/css534/input
./bin/hdfs dfs –put etc/hadoop/*.xml /user/css534/input/
./bin/hdfs dfs –ls /user/css534/input

(6) Run a MapReduce application and check the results
./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar grep input output 'dfs[a-z.]+'
./bin/hdfs dfs –ls /user/css534/output
./bin/hdfs dfs -get /user/css534/output/part-r-00000 ~
cat ~/part-r-00000

(7) Delete /user/css534/output for a next MapReuce run
./bin/hdfs dfs –rm /user/css534/output/*
./bin/hdfs dfs –rmdir /user/css534/output

(8) Compile a MapReduce application, (i.e., WordCount.java)
Copy ~css534/programming/MapReduce/worcount_2.0 to your directory.
cp –r ~css534/programming/MapReduce/wordcount_2.0 .
cd wordcount_2.0
javac –cp `hadoop classpath`:. WordCount.java
jar –cvf wordcount.jar *.class

(9) Delete /user/css534/intput's previous contents and copy wordcount_2.0/intput/* to there.
~/hadoop-2.6.0/bin/hdfs dfs –rm /user/css534/input/*
~/hadoop-2.6.0/bin/hdfs dfs –put intput/* /user/css534/input
~/hadoop-2.6.0/bin/hdfs dfs –ls /user/css534/input

(10) Run WordCount and check the results
~/hadoop-2.6.0/bin/hadoop jar ./wordcount.jar WordCount input output
~/hadoop-2.6.0/bin/hdfs dfs –ls /user/css534/output
~/hadoop-2.6.0/bin/hdfs dfs -get /user/css534/output/part-00000 output
cat output/*

(11) Delete /user/css534/output for a next MapReduce run
~/hadoop-2.6.0/bin/hdfs dfs –rm /user/css534/output/*
~/hadoop-2.6.0/bin/hdfs dfs –rmdir /user/css534/output

(12) Stop HDFS
~/hadoop-2.6.0/sbin/stop-dfs.sh

(13) If you need to reformat your file system.
Delete all your files under /tmp at each machine, (i.e., uw1-320-10, 11, 12, and 13 in this example) Thereafter type:

cd ~/hadoop-2.6.0
./bin/hdfs namenode –format