

Set up each node using the single-node setup guide

Moving to multinode

```
$ /bin/stop-all.sh
```

hduser on the master must be able to connect to itself (ssh master) and also to hduser on the slave via a password-less ssh login

add the huser@master public ssh key to authorized_keys of hduser@slave in the hduser@slave's \$HOME/.ssh/authorized_keys

```
$ ssh-copy-id -i $HOME/.ssh/id_rsa.pub hduser@slave
```

test ssh setup:

connect to slave using master

```
$ ssh master
```

```
$ ssh slave
```

Configuration

conf/masters (master only) defines on which machines Hadoop will start secondary NameNodes in our multi-node cluster

The primary NameNode and the JobTracker will always be the machines on which you run the bin/start-dfs.sh and bin/start-mapred.sh scripts,

on master, update conf/masters

master

This conf/slaves file lists the hosts, one per line, where the Hadoop slave daemons (DataNodes and TaskTrackers) will be run

On master, update conf/slaves that it looks like this:

master

slave

You have to change the configuration files conf/core-site.xml, conf/mapred-site.xml and conf/hdfs-site.xml on ALL machines as follows.

```
<!-- In: conf/core-site.xml -->
```

```
<property>
```

```
<name>fs.default.name</name>
```

```
<value>hdfs://master: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>
```

```
<!-- In: conf/mapred-site.xml -->
```

```
<property>
```

```
<name>mapred.job.tracker</name>
```

```
<value>master: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>
```

```
<!-- In: conf/hdfs-site.xml -->
```

```
<property>
  <name>dfs.replication</name>
  <value>2</value>
  <description>Default block replication.
  The actual number of replications can be specified when the file is created.
  The default is used if replication is not specified in create time.
</description>
</property>
<property>
  <name>dfs.name.dir</name>
  <value>/home/hduser/hadoop/tmp/dfs/name</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hduser/hadoop/tmp/dfs/data</value>
</property>
```

format namenode on NameNode (master)
\$ bin/hadoop namenode -format

Starting the multi-node cluster
run bin/start-dfs.sh on the machine that you want NameNode to run on,
this will bring up HDFS with NameNode on the machine this script is ran and DataNodes on
machines listed in conf/slaves
\$ bin/start-dfs.sh

examine the success/failure of this command in logs/hadoop-hduser-datanode-slave.log on
slave

```
hduser@master:/usr/local/hadoop$ jps
14799 NameNode
15314 Jps
14880 DataNode
14977 SecondaryNameNode
hduser@master:/usr/local/hadoop$
```

```
hduser@slave:/usr/local/hadoop$ jps
15183 DataNode
15616 Jps
hduser@slave:/usr/local/hadoop$
```

Browse web interface for hdfs at <master ip>:50070

Run the command /bin/start-mapred.sh on the machine you want the JobTracker to run on
This will bring up the MapReduce cluster with the JobTracker running on the machine you ran
the previous command on,
and TaskTrackers on the machines listed in the conf/slaves file.

```
$ bin/start-mapred.sh
```

```
hduser@master:/usr/local/hadoop$ jps
```

```
16017 Jps
```

```
14799 NameNode
```

```
15686 TaskTracker
```

```
14880 DataNode
```

```
15596 JobTracker
```

```
14977 SecondaryNameNode
```

```
hduser@master:/usr/local/hadoop$
```

```
hduser@slave:/usr/local/hadoop$ jps
```

```
15183 DataNode
```

```
15897 TaskTracker
```

```
16284 Jps
```

```
hduser@slave:/usr/local/hadoop$
```

web interface at <master ip>:50030

Stopping the cluster

if you get an error that says that tasktracker cannot start because the port is current used, head to mapred-site.xml and change the jobtracker ports from 54310 to something else, like 5