**Single node performance test**

For performance test I have put together a few sample files by concatenating textbooks from project [gutenberg](http://www.gutenberg.org/) and run them in the same manner as the sample test above.

Result:

|  |  |  |
| --- | --- | --- |
| File | Size | Wordcount execution time (mm:ss) |
| smallfile.txt | 2MB | 2:17 |
| mediumfile.txt | 35MB | 9:19 |

[Download](http://www.widriksson.com/wp-content/uploads/2014/10/hadoop_sample_txtfiles.tar.gz) sample text files for performance test.

ATTENZIONE! CON FILE TROPPO GRANDI IL RPi VA OUT OF MEMORY

# Hadoop Raspberry Pi Cluster Setup

## Prepare Node1 for cloning

Since we will make a clone of node1 later the settings made here will be the “base” for all new nodes.

### Edit configuration files

#### /etc/hosts

192.168.0.110 node1

192.168.0.111 node2

192.168.0.112 node3

In a more serious setup you should use real DNS to setup name lookup, however to make it easy we will just go with the hosts file.

#### /opt/hadoop/conf/masters

node1

Note: conf/masters file actually tells which node that is the Secondary NameNode. Node1 will become NameNode when we start the NameNode service on that machine.

In /opt/hadoop/conf edit the following configuration files and change from localhost to node1:

#### **core-site.xml**

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/hdfs/tmp</value>

</property>

<property>

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

<value>hdfs://**node1**:54310</value>

</property>

</configuration>

#### **mapred-site.xml**

<configuration>

  <property>

    <name>mapred.job.tracker</name>

    <value>**node1**:54311</value>

  </property>

</configuration>

### Wipe HDFS

Note: In the next step we will completely wipte out the current hdfs storage – all files and data that you have used in hdfs will be lost. When you format the namenode there is also an issue causing the error message: Incompatible namespaceIDs in path/to/hdfs. This can happen when starting/doing file operations on the datanode after the namenode has been formatted. This issue is explained more in detail[*here.*](http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-multi-node-cluster/#javaioioexception-incompatible-namespaceids)

rm -rf /hdfs/tmp/\*

Later on we will format the namenode but we do this to ensure the hdfs filesystem is clean on all the nodes.

## Clone Node1 and setup slaves

Clone the SD Card of node1 to the other SD cards you plan to use for the other nodes. There are various programs that can do this i used [Win32DiskImager](http://sourceforge.net/projects/win32diskimager/).

For each cloned node make sure to:

* Change hostame in /etc/hostname
* Change IP Adress in /etc/network/interfaces
* Restart the Pi.

## Configure Node1

/opt/hadoop/conf/slaves

node1

node2

node3

Note: The masters and slaves configuration files are only read by the hadoop start/stop scripts such as: start-all.sh, start-dfs.sh and start-mapred.sh.

On node1, ensure you can reach node2 and node3 from ssh as hduser without need to enter password. If this does not work: copy /home/hduser/.ssh/id\_rsa.pub on node1 to /home/hduser/.ssh/authorized\_keys on the node that you try to connect to.

su hduser

ssh node1

exit

ssh node2

exit

ssh node3

exit

Enter Yes when you get the “Host key verification failed message”.

## Format hdfs and start services

On **node1**:

hadoop namenode -format

/opt/hadoop/bin/start-dfs.sh

/opt/hadoop/bin/start-mapred.sh

## Verify that daemons are running correctly

On node1:

jps

3729 SecondaryNameNode

4003 Jps

3607 DataNode

3943 TaskTracker

3819 JobTracker

3487 NameNode

On the other nodes:

jps

2307 TaskTracker

2227 DataNode

2363 Jps

Note: If you have issues you can examine the logfiles /opt/hadoop/logs or you can try to start each service manually on the node that is failing for example:

On node1:hadoop namenodehadoop datanode

You may now also try to access hadoop from the web interface to see which nodes that are active and other statistics:

http://node1:50030  
http://node1:50070