

# Big Data - Concepts

23 December 2023 09:00

## Final Steps for implementing two node Hadoop / HDFS clusters

Current setup nodes as follows : mainserver1: 192.168.56.100 slave1: 192.168.56.101 slave2:192.168.56.102

### Perform the following activity on all the nodes

# nano /etc/hosts

```
hadoop@mainserver1:~$ cat /etc/hosts
127.0.0.1 localhost
#127.0.1.1 mainserver1

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters

192.168.56.100 mainserver1
192.168.56.101 slave1
192.168.56.102 slave2
hadoop@mainserver1:~$
```

Hash out the line containing the name of the node against 127.0.0.1 only  
Do not comment out line containing localhost nor edit anything

```
hadoop@slave2:~$ ls
hadoop-2.4.1.tar.gz jdk-7u80-linux-x64.tar.gz
hadoop@slave2:~$ tar -zxf jdk-7u80-linux-x64.tar.gz
hadoop@slave2:~$ tar -zxf hadoop-2.4.1.tar.gz
hadoop@slave2:~$ ls -l
total 285356
drwxr-xr-x 9 hadoop hadoop 4096 Jun 21 2014 hadoop-2.4.1
-rw-rw-r-- 1 hadoop hadoop 138656756 Dec 22 04:22 hadoop-2.4.1.tar.gz
drwxr-xr-x 8 hadoop hadoop 4096 Apr 11 2015 jdk1.7.0_80
-rw-rw-r-- 1 hadoop hadoop 153530841 Dec 22 04:22 jdk-7u80-linux-x64.tar.gz
hadoop@slave2:~$ su
Password:
root@slave2:/home/hadoop# mv hadoop-2.4.1 /usr/local/hadoop
root@slave2:/home/hadoop# mv jdk1.7.0_80 /usr/local/
root@slave2:/home/hadoop# chown -R hadoop.hadoop /usr/local/hadoop
root@slave2:/home/hadoop# chown -R root.root /usr/local/jdk1.7.0_80/
root@slave2:/home/hadoop# exit
exit
hadoop@slave2:~$
```

# nano /etc/bash.bashrc

<add the following lines to the end of the file >

```
export JAVA_HOME=/usr/local/jdk1.7.0_80
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_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_INSTALL=$HADOOP_HOME
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

<save and exit>

\$ su

# mkdir -p /opt/hadoop

# chown -R hadoop.hadoop /opt/hadoop

# exit

Logout and login as hadoop in both the nodes and run the following command

```
# java -version
# hadoop version
```

we should be able to get the following output after executing the above command on both the nodes

```
hadoop@slave1:~$
hadoop@slave1:~$ java -version
java version "1.7.0_80"
Java(TM) SE Runtime Environment (build 1.7.0_80-b15)
Java HotSpot(TM) 64-Bit Server VM (build 24.80-b11, mixed mode)
hadoop@slave1:~$ hadoop version
Hadoop 2.4.1
Subversion http://svn.apache.org/repos/asf/hadoop/common -r 1604318
Compiled by jenkins on 2014-06-21T05:43Z
Compiled with protoc 2.5.0
From source with checksum bb7ac0a3c73dc131f4844b873c74b630
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-2.4.1.jar
```

**Edit hadoop-env.sh and add the following entry - on all the nodes**

```
# cd $HADOOP_HOME/etc/hadoop
# nano hadoop-env.sh
# The java implementation to use.
export JAVA_HOME=/usr/local/jdk1.7.0_80
```

Now execute the following being in /usr/local/hadoop/etc/hadoop directory - note you should have logged in as hadoop user only

```
$ cd $HADOOP_HOME/etc
```

```
$ nano core-site.xml
```

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://mainserver1:9000</value>
  </property>
</configuration>
```

<save and exit>

```
$ nano hdfs-site.xml
```

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>

  <property>
    <name>dfs.name.dir</name>
    <value>file:///opt/hadoop/hadoopinfra/hdfs/namenode</value>
  </property>

  <property>
    <name>dfs.data.dir</name>
    <value>file:///opt/hadoop/hadoopinfra/hdfs/datanode</value>
  </property>
</configuration>
```

<save and exit>

```
$ nano yarn-site.xml
```

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

  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>

  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>mainserver1</value>
  </property>
```

```

<property>
  <name>yarn.resourcemanager.address</name>
  <value>mainserver1:8032</value>
</property>
</configuration>
<save and exit>

```

```
$ cp mapred-site.xml.template mapred-site.xml
```

```
$ nano mapred-site.xml
```

```

<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>mainserver1:9001</value>
  </property>
</configuration>
<save and exit>

```

```
# nano slaves
```

```
slave1
```

```
slave2
```

```
<save and exit>
```

```
# nano masters { add the following entry only in mainserver1 and in slaves file should be there but it should be empty }
```

```
mainserver1
```

```
<save and exit>
```

```
$ hdfs namenode -format
```

```
$ start-dfs.sh
```

```
$ start-yarn.sh
```

```

192.168.56.100 - PuTTY
hadoop@mainserver1:~$ jps
6471 SecondaryNameNode
6229 NameNode
6597 ResourceManager
7191 Jps
hadoop@mainserver1:~$ ss -ant
State      Recv-Q    Send-Q      Local Address:Port      Peer Address:Port      Process
LISTEN     0          128         127.0.0.53%lo:53        0.0.0.0:*               *
LISTEN     0          128         0.0.0.0:50070          0.0.0.0:*               *
LISTEN     0          128         0.0.0.0:22            0.0.0.0:*               *
LISTEN     0          128        192.168.56.100:9000    0.0.0.0:*               *
LISTEN     0          128         0.0.0.0:50090        0.0.0.0:*               *
ESTAB      0          0        192.168.56.100:9000    192.168.56.101:53062    *
ESTAB      0          0        192.168.56.100:9000    192.168.56.102:45486    *
TIME-WAIT  0          0        192.168.56.100:52754    192.168.56.100:9000    *
ESTAB      0          64        192.168.56.100:22      192.168.56.1:38068      *
ESTAB      0          0        192.168.56.100:22      192.168.56.1:50483      *
ESTAB      0          0        192.168.56.100:22      192.168.56.1:51040      *
LISTEN     0          128          [::]:22                [::]:*                  *
LISTEN     0          128    [::ffff:192.168.56.100]:8088    *:8088                   *
LISTEN     0          128    [::ffff:192.168.56.100]:8030    *:8030                   *
LISTEN     0          128    [::ffff:192.168.56.100]:8031    *:8031                   *
LISTEN     0          128    [::ffff:192.168.56.100]:8032    *:8032                   *
LISTEN     0          128    [::ffff:192.168.56.100]:8033    *:8033                   *
ESTAB      0          0    [::ffff:192.168.56.100]:8031    [::ffff:192.168.56.101]:52760    *
ESTAB      0          0    [::ffff:192.168.56.100]:8031    [::ffff:192.168.56.102]:50816    *
hadoop@mainserver1:~$ █

```

```
hadoop@slave1:~$ jps
5556 NodeManager
5830 Jps
5375 DataNode
hadoop@slave1:~$ ss -ant
State Recv-Q Send-Q Local Address:Port Peer Address:Port Process
LISTEN 0 4096 127.0.0.53:10:53 0.0.0.0:*
LISTEN 0 128 0.0.0.0:22 0.0.0.0:*
LISTEN 0 50 0.0.0.0:50010 0.0.0.0:*
LISTEN 0 128 0.0.0.0:50075 0.0.0.0:*
LISTEN 0 128 0.0.0.0:50020 0.0.0.0:*
ESTAB 0 0 192.168.56.101:53062 192.168.56.100:9000
ESTAB 0 64 192.168.56.101:22 192.168.56.1:38073
ESTAB 0 0 192.168.56.101:22 192.168.56.1:50521
LISTEN 0 128 *:8040 *:*
LISTEN 0 128 *:8042 *:*
LISTEN 0 128 *:45685 *:*
LISTEN 0 128 [::]:22 [::]:*
LISTEN 0 50 *:13562 *:*
ESTAB 0 0 [::ffff:192.168.56.101]:52760 [::ffff:192.168.56.100]:8031
hadoop@slave1:~$
```

```
hadoop@slave2:~$ jps
4688 DataNode
5162 Jps
4869 NodeManager
hadoop@slave2:~$ ss -ant
State Recv-Q Send-Q Local Address:Port Peer Address:Port Process
LISTEN 0 4096 127.0.0.53:10:53 0.0.0.0:*
LISTEN 0 128 0.0.0.0:22 0.0.0.0:*
LISTEN 0 50 0.0.0.0:50010 0.0.0.0:*
LISTEN 0 128 0.0.0.0:50075 0.0.0.0:*
LISTEN 0 128 0.0.0.0:50020 0.0.0.0:*
ESTAB 0 0 192.168.56.102:22 192.168.56.1:50551
ESTAB 0 0 192.168.56.102:45486 192.168.56.100:9000
ESTAB 0 64 192.168.56.102:22 192.168.56.1:38190
LISTEN 0 128 *:8040 *:*
LISTEN 0 128 *:8042 *:*
LISTEN 0 128 *:35405 *:*
LISTEN 0 128 [::]:22 [::]:*
LISTEN 0 50 *:13562 *:*
ESTAB 0 0 [::ffff:192.168.56.102]:50816 [::ffff:192.168.56.100]:8031
hadoop@slave2:~$
```

Cluster

About Nodes ApplicationsNEW NEW SAVING SUBMITTED ACCEPTED RUNNING FINISHED FAILED KILLED SchedulerTools

### Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	16 GB	0 B	2	0	0	0	0

Show 20 entries

Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail	Version
/default-rack	RUNNING	slave2:35405	slave2:8042	23-Dec-2023 07:51:02		0	0 B	8 GB	2.4.1
/default-rack	RUNNING	slave1:45685	slave1:8042	23-Dec-2023 07:51:02		0	0 B	8 GB	2.4.1

Showing 1 to 2 of 2 entries

About Apache Hadoop

## Overview 'mainserver1:9000' (active)

Started:	Sat Dec 23 07:49:25 UTC 2023
Version:	2.4.1, r1604318
Compiled:	2014-06-21T05:43Z by Jenkins from branch-2.4.1
Cluster ID:	CID-6ebddadc-7d7f-44bd-8f57-b2f6775135cc
Block Pool ID:	BP-200813437-192.168.56.100-1703317718255

## Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 32.7 MB of 43.78 MB Heap Memory. Max Heap Memory is 966.69 MB.

Non Heap Memory used 27.2 MB of 28.5 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	56.75 GB
DFS Used:	48 KB
Non DFS Used:	13.64 GB
DFS Remaining:	43.1 GB
DFS Used%:	0%
DFS Remaining%:	75.96%
Block Pool Used:	48 KB
Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
<a href="#">Live Nodes</a>	2 (Decommissioned: 0)
<a href="#">Dead Nodes</a>	0 (Decommissioned: 0)
<a href="#">Decommissioning Nodes</a>	0
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0

## NameNode Journal Status

Current transaction ID: 3

Journal Manager	State
FileJournalManager(root=/opt/hadoop/hadoopinfra/hdfs/namenode)	EditLogFileOutputStream(/opt/hadoop/hadoopinfra/hdfs/namenode/current/edits_inprogress_0000000000000000003)

## NameNode Storage

Storage Directory	Type	State
/opt/hadoop/hadoopinfra/hdfs/namenode	IMAGE_AND_EDITS	Active

# Datanode Information

## In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
slave1 (192.168.56.101:50010)	0	In Service	28.37 GB	24 KB	6.78 GB	21.59 GB	0	24 KB (0%)	0	2.4.1
slave2 (192.168.56.102:50010)	2	In Service	28.37 GB	24 KB	6.86 GB	21.51 GB	0	24 KB (0%)	0	2.4.1

## Decomissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks in files under construction
------	--------------	-------------------------	------------------------------	--