**Additions from single node cluster to Multi-node cluster**

# update the hosts file on all the machines

**cd /etc**

**gedit hosts**

**# Sample IP addresses. We can have the IP addresses for all the machines listed here with hostnames.**

10.2.40.202 Hadoop01

10.2.40.203 Hadoop02

10.2.40.204 Hadoop03

10.2.40.205 Hadoop04

10.2.40.206 Hadoop05

10.2.40.207 Hadoop06

**SSH access**

The hduser user on the master (aka [hduser@](mailto:hduser%40master)Hadoop01) must be able to connect:

1. to its own user account on the master - i.e. ssh master in this context.
2. to the hduser user account on the slave (i.e. [hduser@](mailto:hduser%40slave1)Hadoop02/03/04/05/06) via a password-less SSH login.

* Add the [hduser@](mailto:hduser%40master)Hadoop01 public SSH key using the following command
* hduser@Hadoop01:~$ ssh-copy-id -i $HOME/.ssh/id\_rsa.pub hduser@Hadoop02
* hduser@Hadoop01:~$ ssh-copy-id -i $HOME/.ssh/id\_rsa.pub hduser@Hadoop03
* hduser@Hadoop01:~$ ssh-copy-id -i $HOME/.ssh/id\_rsa.pub hduser@Hadoop04
* hduser@Hadoop01:~$ ssh-copy-id -i $HOME/.ssh/id\_rsa.pub hduser@Hadoop05
* hduser@Hadoop01:~$ ssh-copy-id -i $HOME/.ssh/id\_rsa.pub hduser@Hadoop06
* Connect with user hduser from the master to the user account hduser on the slave.

1. From master to master
2. hduser@Hadoop01:~$ ssh Hadoop01
3. From master to all slaves
4. hduser@Hadoop01:~$ ssh Hadoop02
5. hduser@Hadoop01:~$ ssh Hadoop03
6. hduser@Hadoop01:~$ ssh Hadoop04
7. hduser@Hadoop01:~$ ssh Hadoop05
8. hduser@Hadoop01:~$ ssh Hadoop06

**Configuration**

**/usr/local/hadoop/etc/hadoop/masters**

The machine on which bin/start-dfs.sh is running will become the primary NameNode. This file should be updated on all the nodes. Open the masters file in the conf directory

hduser@Hadoop01/Hadoop02/03/04/05/06 :~$ cd /usr/local/hadoop/etc/hadoop/

hduser@Hadoop01/Hadoop02/03/04/05/06 :~$ sudo gedit masters

Add the following line

Hadoop01

**usr/local/hadoop/etc/hadoop/slaves**

This file should be updated on all the nodes as master is also a slave. Open the slaves file in the conf directory

hduser@Hadoop01/Hadoop02/03/04/05/06:$ cd /usr/local/hadoop/etc/hadoop

$ sudo gedit slaves

Add the following lines

Hadoop02

Hadoop03

Hadoop04

Hadoop05

Hadoop06

**CORE-SITE.XML**

Change the fs.default.name parameter (in core-site.xml), which specifies the NameNode (the HDFS master) host and port.

/usr/local/hadoop/etc/hadoop/core-site.xml (ALL machines .ie. Master as well as slave)

<property>

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

<value>hdfs://**Hadoop01**: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>

**mapred-site.xml**

Open this file in the conf directory

hduser@Hadoop01:~$ /usr/local/hadoop/etc/hadoop

hduser@Hadoop01:~$ sudo gedit mapred-site.xml

Change the mapred.job.tracker parameter (in conf/mapred-site.xml), which specifies the JobTracker (MapReduce master) host and port.

/usr/local/hadoop/etc/hadoop/mapred-site.xml (ALL machines)

<property>

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

<value>**Hadoop01**: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>

**hdfs-site.xml**

Open this file in the conf directory

hduser@Hadoop01:~$ /usr/local/hadoop/etc/hadoop

hduser@Hadoop01:~$ sudo gedit hdfs-site.xml

Change the dfs.replication parameter (in hdfs-site.xml) which specifies the default block replication. We have six nodes available, so we set dfs.replication to 2

**hdfs-site.xml (ALL machines)**

Changes to be made

<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>

**Formatting the HDFS filesystem via the NameNode**

Format the cluster’s HDFS file system

hduser@Hadoop01:~$ hadoop namenode -format

**Starting the multi-node cluster**

Starting the cluster is performed in two steps.

1. We begin with starting the HDFS daemons: the NameNode daemon is started on master, and DataNode daemons are started on all slaves (here: master and slaves).
2. Then we start the MapReduce daemons: the JobTracker is started on master, and TaskTracker daemons are started on all slaves (here: master and slave).

Cluster is started by running the command only on master

hduser@Hadoop01:~$ **start-all.sh**

**On the Namenode**

$ **jps**

jps

Namenode

Secondary Namenode

Resource Manager

**On the Datanodes**

$ **jps**

jps

Datanode

Node Manager