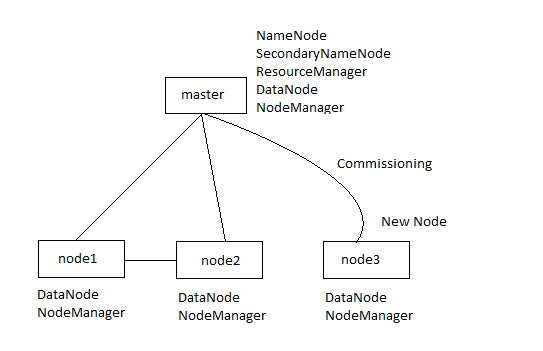
**Adding a New DataNode in the Hadoop Cluster**

In this tutorial, I will show you how to add new nodes to an existing HDFS cluster without restarting the whole cluster, and how to force HDFS to rebalance after the addition of new nodes.



**Prerequisites:**

1. Multi-node Hadoop cluster
2. A new datanode

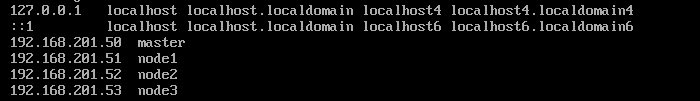
If you don’t have Hadoop VMs, you can access the following link to download.

**https://drive.google.com/open?id=1WYIOJfupaq2X411KWEROJF1anAD5mfYQ**

(Paste the link on web browser to download these files in **Commissioning\_Cluster\_3Nodes** folder including **master.zip, node1.zip, node2.zip**. VMs User/Password: root/123456 or hadoopuser/123456)

**Setup new datanode (node3):**

1. Clone an existing DataNode (See **01\_Multi-node installation** for the instruction)
2. Configure IP, hostname, and modify **/etc/hosts** and then update for other machines (**master, node1, node2**)



1. Delete data folder

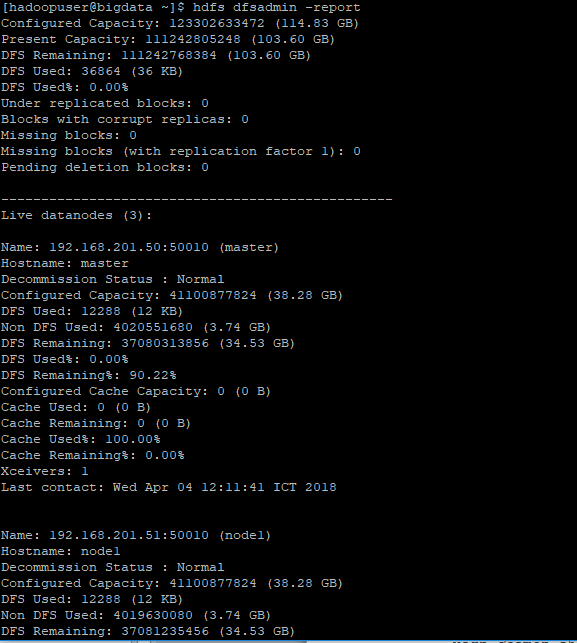
*rm -rf /opt/niit/hadoop\_store/hdfs/namenode/\**

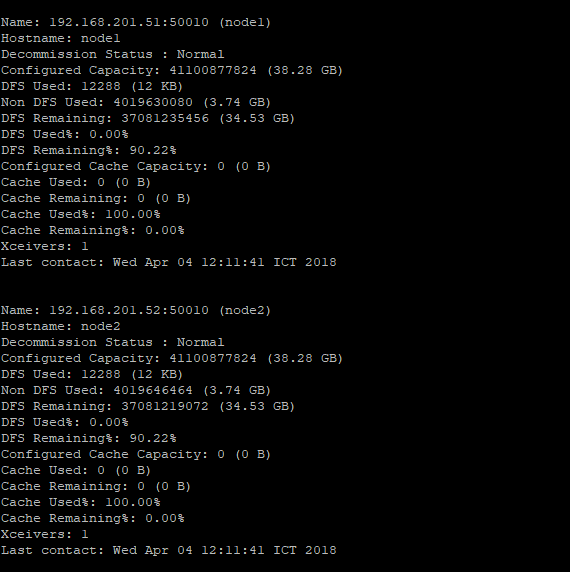
*rm -rf /opt/niit/hadoop\_store/hdfs/datanode/\**

**Check Hadoop cluster health:**

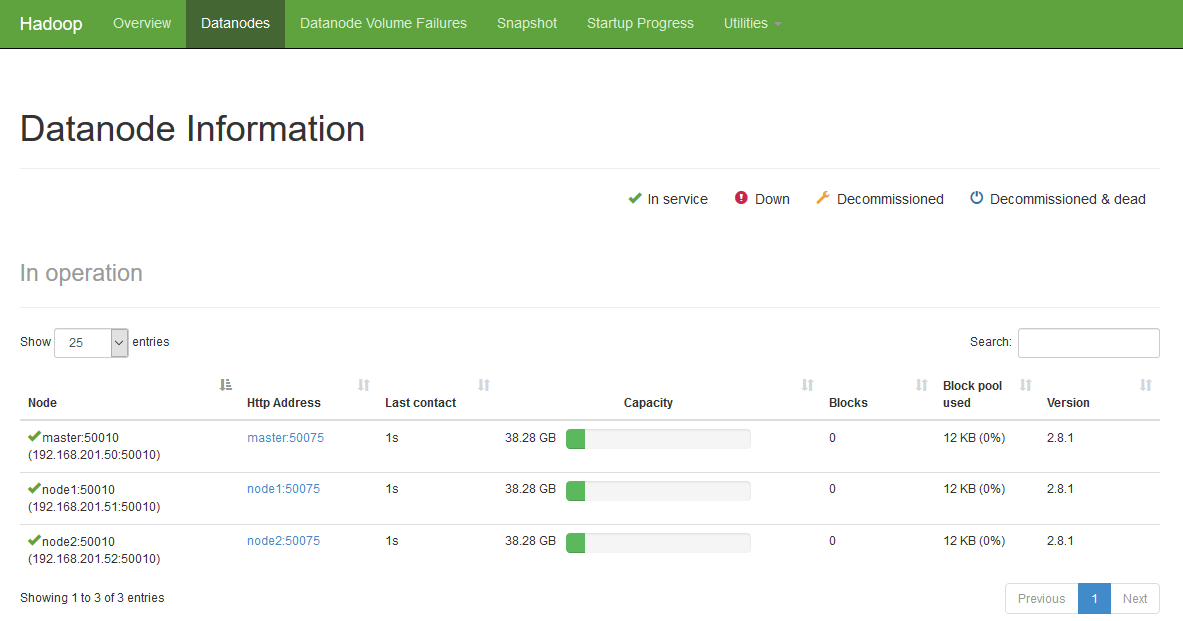
Currently, we have three live datanodes in our Hadoop cluster system. Check Hadoop cluster report with the following command.

*hdfs dfsadmin –report*





Or access the webpage **http://192.168.201.50:8080** to view information.

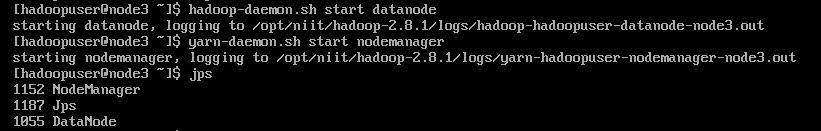


**Commissioning of Datanodes:**

1. Start **DataNode** and **NodeManager** on node3

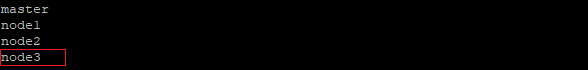
*hadoop-daemon.sh start datanode*

*yarn-daemon.sh start nodemanager*

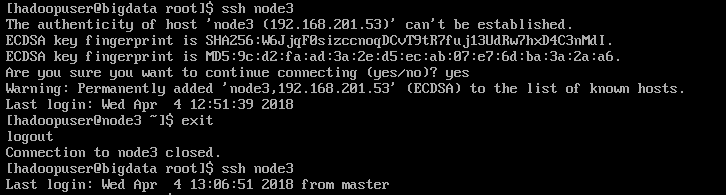


1. **On master,** update **$HADOOP\_HOME/etc/hadoop/slaves** file:

Add new node to the bottom of file



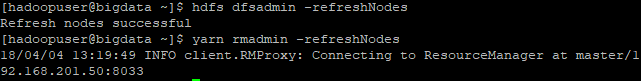
1. Check ssh from master to node3 without any prompt for password.



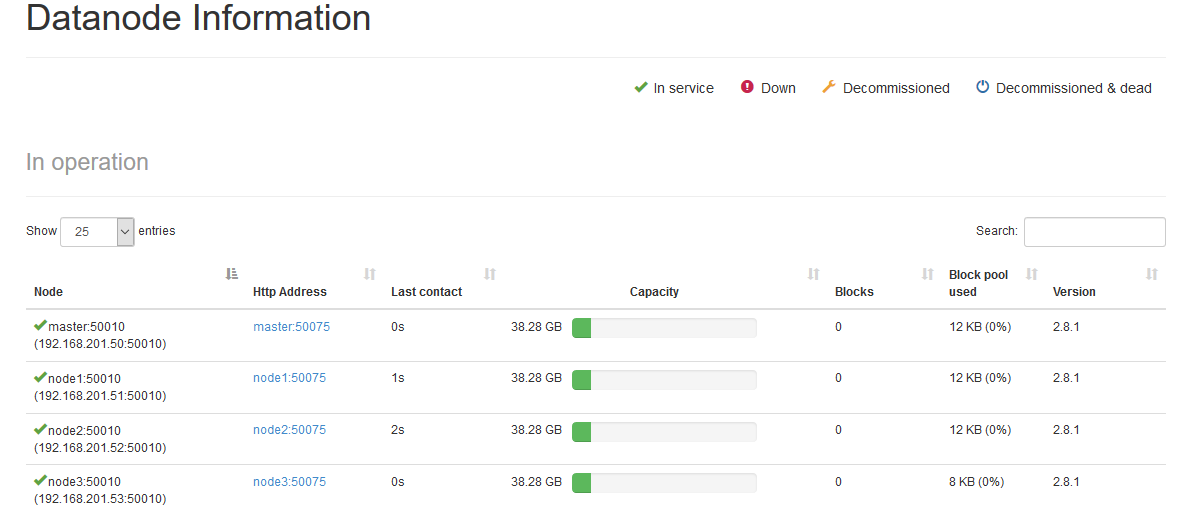
1. Go to master and perform a refresh

*hdfs dfsadmin –refreshNodes*

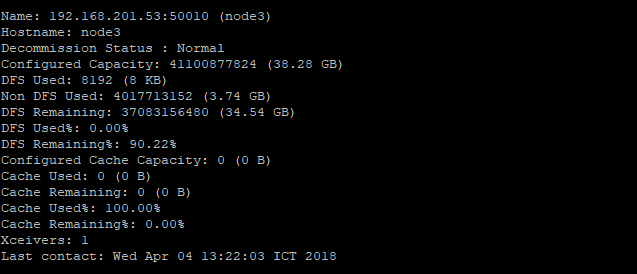
*yarn rmadmin –refreshNodes*



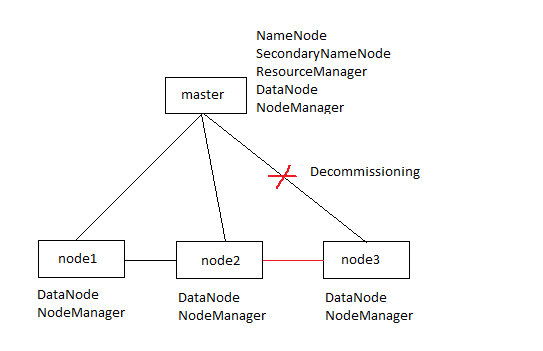
1. Check result on web **http://192.168.201.50:8080**



Check hadoop report: *hdfs dfsadmin –report*



**Decommissioning of Datanodes:**



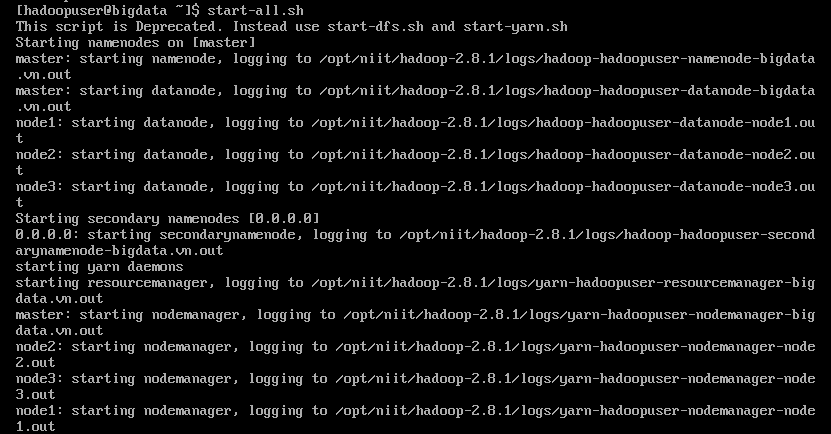
If you don’t have Hadoop VMs, you can access the following link to download.

**https://drive.google.com/open?id=1WYIOJfupaq2X411KWEROJF1anAD5mfYQ**

(Paste the link on web browser to download these files in **Decommissioning\_Cluster\_4Nodes** folder including **master.zip, node1.zip, node2.zip, node3.zip**. VMs User/Password: root/123456 or hadoopuser/123456)

1. Check Hadoop daemons on all nodes. If they are not running, we go to **master** and use the following command to start.

start-all.sh



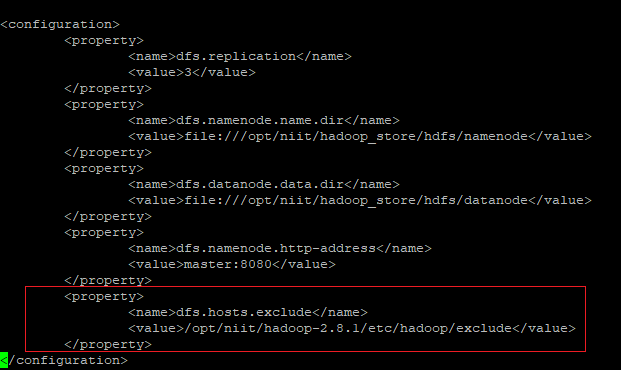
1. Add the following lines to **hdfs-site.xml** file

*<property>*

*<name>dfs.hosts.exclude</name>*

*<value>/opt/niit/hadoop-2.8.1/etc/hadoop/exclude</value>*

*</property>*



1. Create **/opt/niit/hadoop-2.8.1/etc/hadoop/exclude** file with the following line.

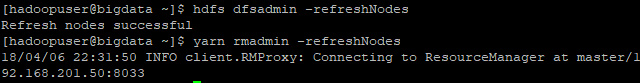
*node3*



1. Refresh. On **master,** enter the following commands.

*hdfs dfsadmin -refreshNodes*

*yarn rmadmin –refreshNodes*



1. Check Hadoop nodes status

*hdfs dfsadmin –report*

