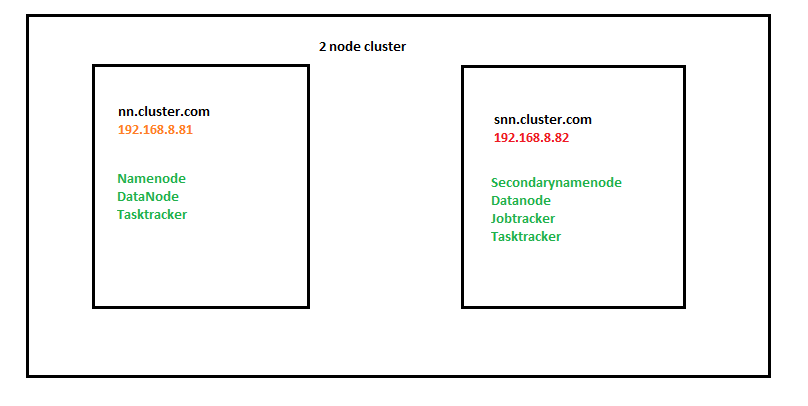
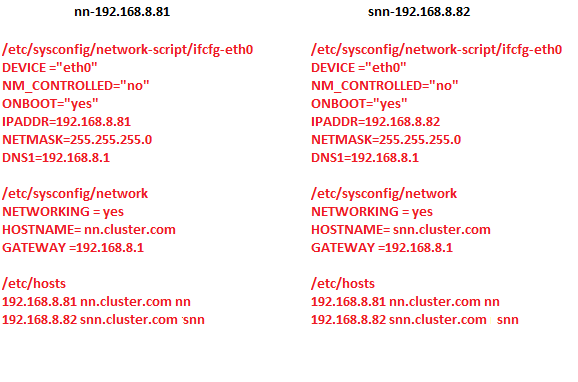
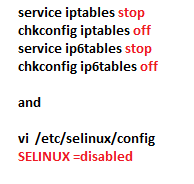
**Project 1**

**Cluster planning**

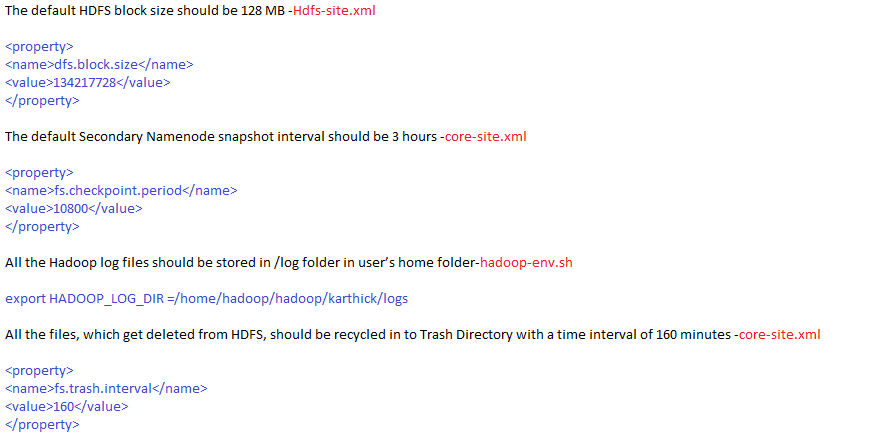


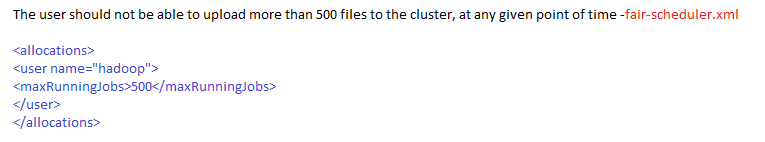


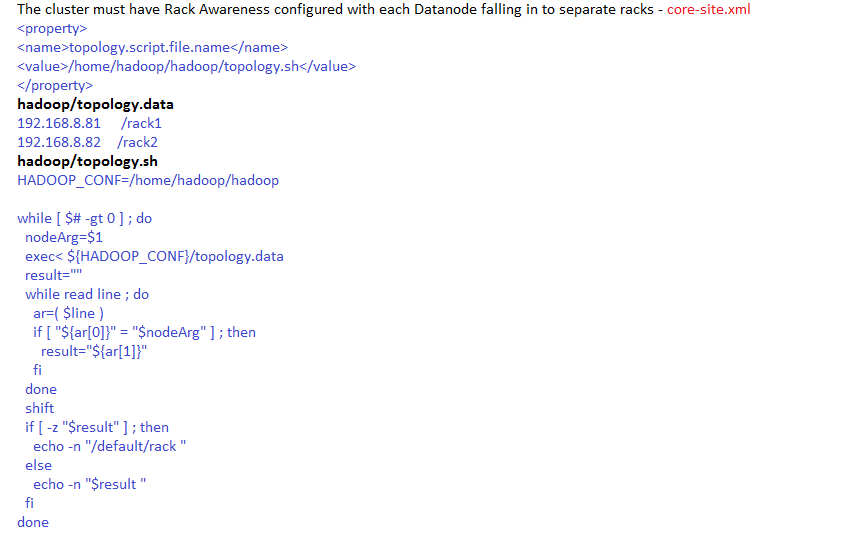
Have to make sure



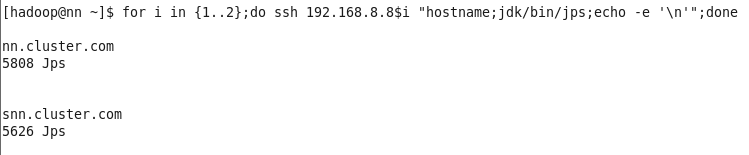
**Configuration**

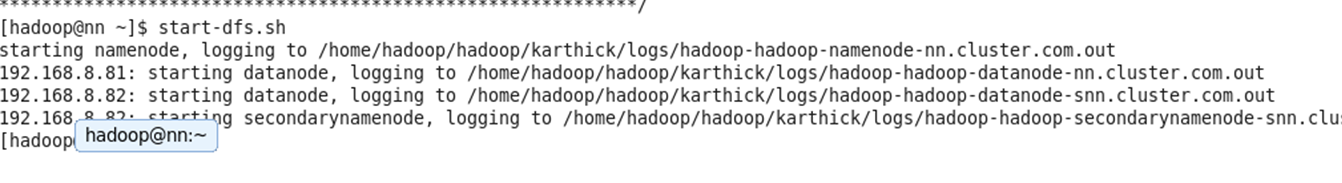


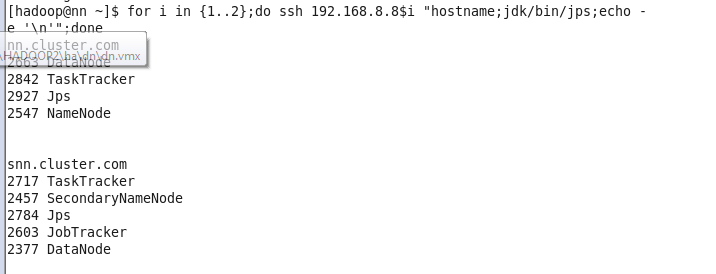


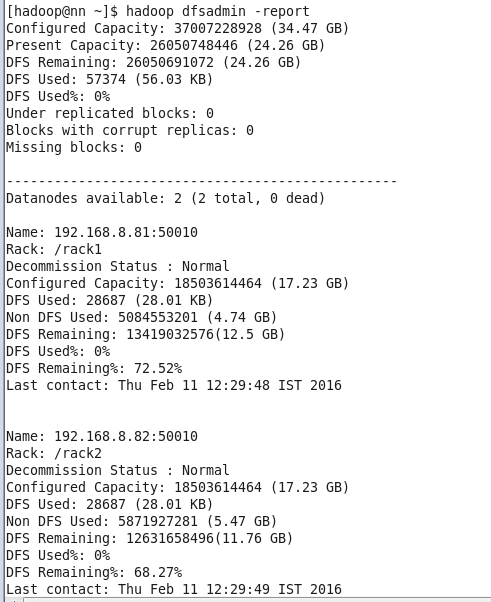


**Screen shot for the 2 node cluster working fine**









Once completed, pen down the step by step process to upgrade this cluster from Hadoop 1.2.1 to Hadoop 2.6.0

Steps to upgrade:

1. Stop all the currently running jobs and stop-mapred.sh to stop mapreduce processes.

2. Take the backup of metadata with following commands:

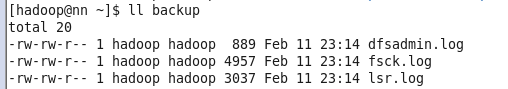
a. mkdir backup

b. hadoop fsck / -blocks -files -locations -racks > backup/fsck.log

c. hadoop fs -lsr / > backup/lsr.log

d. hadoop dfsadmin -report > backup/dfsadmin.log

e. cp -rv data/nn > backup/



3. (optional) - Take the backup of the HDFS data.

4. stop-dfs.sh - this will stop all the processes in the cluster.

5. Install hadoop version 2.

a. in version 2: same properties in core-site.xml/hdfs-site.xml/slaves/hadoop-env.sh need to be added with the foll. exceptions:

b. fs.default.name becomes - fs.defaultFS

c. dfs.name.dir becomes dfs.namenode.name.dir

d. dfs.data.dir becomes dfs.datanode.data.dir

e. no masters file for the snn, instead of that add the following property in hdfs-site.xml in NN.

dfs.namenode.secondary.http-addres = <ip of snn>:50090

dfs.namenode.checkpoint.dir = <snn checkpoint dir>

f. Slaves and hadoop-env.sh files will remain same

g. Also, change the .bash\_profile file accordingly.





Note: We are not adding anything for the mapreduce as it is a newer feature in HADOOP 2.

6. Upgrade the namenode with the following command:

hadoop-daemon.sh start namenode -ugprade

start the datanode and secondary namenode with the following commands:

hadoop-daemon.sh start datanode

hadoop-daemon.sh start secondarynamenode

7. If there are any issues in the upgrade, follwing steps needs to be used to rollback:

a. stop-dfs.sh

b. remove the nn metadata directory.

c. copy the old nn metadata directory to the same location.

d. start-dfs.sh

8. Finalize the upgrade, if everthing is successful.

hdfs dfsadmin -finalizeUpgrade

Note: you won't be able to rollback after this command.



Layout a plan to expand this cluster to accommodate 50 TB of data. The plan should include options to add nodes seamlessly, without formatting. Consider a MOM growth of 20 TB for the cluster. List out the sample configurations for Datanodes as well.

