Remove a Datanode
==========

1. Decommission a host gracefully

```
<property>
      <name>dfs.hosts.exclude</name>
      <value>/home/hadoop/excludes</value>
      <final>true</final>
</property>
```

Similarly for Jobtracker.

```
<property>
      <name>mapred.hosts.exclude</name>
      <value>/home/hadoop/excludes</value>
      <final>true</final>
</property>
```

mapred.hosts.exclude in mapred-site.xml

2. Add the FQDN to the exclude file and refresh the Configurations

3. Update for the Namenode

```
>hadoop dfsadmin -refreshNodes
```

4. Update for Jobtracker

```
>hadoop mradmin -refreshNodes
```

Add Hosts:

1. dfs.hosts in the hdfs-site.xml, mapred.hosts

==========

Cluster Balancing

```
>hadoop balancer -threshold 40
```

==========

Add Disk Space to a datanode

How do you add storage to cluster

```
<property>
      <name>dfs.hosts</name>
      <value>/home/hadoop/include</value>
      <final>true</final>
</property>
```

Hadoop Backup and Recovery
==========

1. Secondary namenode checkpointing

If you want to explicitly specify the file to be used by the namenode

```
      hadoop-daemons.sh --hosts masters start secondarynamenode

      hdfs secondarynamenode -checkpoint force
```

2. hadoop namenode -importCheckpoint

```
<property>
      <name>fs.checkpoint.dir</name>
      <value>/data/new</value>
</property>
```

3. Save NameSpace

```
hadoop dfsadmin -safemode enter

hadoop dfsadmin -saveNamespace
```

Remember it updates under the Namespace directory.

4. Metadata Save

```
hdfs dfsadmin -metasave filename.txt
```

5. Can do a detailed view of the namespace (above 0.21)

```
hdfs oiv -i /data/namenode/current/fsimage -o fsimage.txt
```

Hadoop Upgrade
==========

1. hadoop dfsadmin -upgradeProgress status

2. Stop all client applications running on the MapReduce cluster.

3. Perform a filesystem check
   hadoop fsck / -files -blocks -locations > dfs-v-old-fsck-1.log

4. Save a complete listing of the HDFS namespace to a local file
   hadoop dfs -lsr / > dfs-v-old-lsr-1.log

5. Create a list of DataNodes participating in the cluster:
   hadoop dfsadmin -report > dfs-v-old-report-1.log

6. Optionally backup HDFS data

7. Upgrade process:
   Point to the new directory, update environment variables.

8. hadoop-daemon.sh start namenode -upgrade

9. hadoop dfsadmin -upgradeProgress status

10. Now start the datanode, after pointing to the new hadoop directory

11. hadoop dfsadmin -safemode get

12. hadoop dfsadmin –finalizeUpgrade

Applying Quota
----

hadoop dfsadmin -setSpaceQuota 1m

dfsadmin -setQuota <N> <directory>

dfsadmin -clrQuota <directory>

dfsadmin -setSpaceQuota <N>

dfsadmin -clrSpaceQuota

Distcp
======

hadoop distcp hdfs://nn1:8020/foo/bar hdfs://nn2:8020/bar/foo

hdfs://nn1:8020/foo/a hdfs://nn1:8020/foo/b

hadoop distcp hdfs://nn1.cluster1.com:9000/jobtracker
hdfs://nn1.cluster1.com:9000/newtracker