***High Availability setup***

1. Download zookeeper from the below url

**wget** https**://**archive.apache.org**/**dist**/**zookeeper**/**zookeeper-3.4.6**/**zookeeper-3.4.6.tar.gz

**tar** –xvf zookeeper-3.4.6.tar.gz

**mv** zookeeper-3.4.6 **/**usr**/**local**/**zookeeper

2. update Zookeeper home in bashrc

**export** ZOOKEEPER\_HOME**=/**usr**/**local**/**zookeeper**/**

**export** PATH**=$PATH:$ZOOKEEPER\_HOME/**bin

source .bashrc

3. Edit Core-site.xml

<configuration>

<property>

<name>**hadoop.tmp.dir**</name>

<value>**/usr/local/hadoop\_store/tmp**</value>

</property>

<property>

<name>**fs.defaultFS**</name>

<value>**hdfs://ha-cluster**</value>

</property>

<property>

<name>**dfs.journalnode.edits.dir**</name>

<value>**/usr/local/hadoop\_store/jn**</value>

</property>

</configuration>

4. Edit Hdfs-site.xml

<configuration>

<property>

<name>**dfs.replication**</name>

<value>**1**</value>

</property>

<property>

<name>**dfs.replication**</name>

<value>**1**</value>

</property>

<property>

<name>**dfs.namenode.name.dir**</name>

<value>**file:/usr/local/hadoop\_store/hdfs/namenode**</value>

</property>

<property>

<name>**dfs.datanode.data.dir**</name>

<value>**file:/usr/local/hadoop\_store/hdfs/datanode**</value>

</property>

<property>

<name>**dfs.namenode.checkpoint.dir**</name>

<value>**file:/usr/local/hadoop\_store/hdfs/secondarynamenode**</value>

</property>

<property>

<name>**dfs.namenode.checkpoint.period**</name>

<value>**3600**</value>

</property>

<property>

<name>**dfs.nameservices**</name>

<value>**ha-cluster**</value>

</property>

<property>

<name>**dfs.ha.namenodes.ha-cluster**</name>

<value>**namenode,datanode2**</value>

</property>

<property>

<name>**dfs.namenode.rpc-address.ha-cluster.namenode**</name>

<value>**namenode:9000**</value>

</property>

<property>

<name>**dfs.namenode.rpc-address.ha-cluster.datanode2**</name>

<value>**datanode2:9000**</value>

</property>

<property>

<name>**dfs.namenode.http-address.ha-cluster.namenode**</name>

<value>**namenode:50070**</value>

</property>

<property>

<name>**dfs.namenode.http-address.ha-cluster.datanode2**</name>

<value>**datanode2:50070**</value>

</property>

<property>

<name>**dfs.namenode.shared.edits.dir**</name>

<value>**qjournal://namenode:8485;datanode2:8485;datanode1:8485/ha-cluster**</value>

</property>

<property>

<name>**dfs.client.failover.proxy.provider.ha-cluster**</name>

<value>**org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider**</value>

</property>

<property>

<name>**dfs.ha.automatic-failover.enabled**</name>

<value>**true**</value>

</property>

<property>

<name>**ha.zookeeper.quorum**</name>

<value>**namenode:2181,datanode1:2181,datanode2:2181**</value>

</property>

<property>

<name>**dfs.ha.fencing.methods**</name>

<value>**sshfence**</value>

</property>

<property>

<name>**dfs.ha.fencing.ssh.private-key-files**</name>

<value>**/home/hduser/.ssh/id\_rsa**</value>

</property>

</configuration>

5. Edit zookeeper related details in zookeper config files

**cd** **/**usr**/**local**/**zookeeper**/**

**cd** conf

**cp** zoo\_sample.cfg zoo.cfg

**vi** zoo.cfg

dataDir=/usr/local/hadoop\_store/zookeeper/

#note we have to create above directory

Server.1=namenode:2888:3888

Server.2=datanode1:2888:3888

Server.3=datanode2:2888:3888

6. copy all resourcses to all nodes

bashrc,core-site.xml,hdfs-site.xml,zookeeper

scp -r zookeeper/ hduser@datanode1:/usr/local/

scp -r hdfs-site.xml hduser@datanode1:/usr/local/hadoop/etc/hadoop

etc..

7. Goto zookeeper dataDir

//namenode

vi myid

1

//standby

vi myid

2

//datanode

vi myid

3

8. In all nodes start the journal nodes run the below cmd to start

hadoop-daemon.sh start journalnode

9. If it is fresh instalation format the namenode (skip for existing nstalation)

hdfs namenode -format

11. start namenode

hadoop- daemon.sh start namenode

12. In stand by namenode run the belwo command to copy the metadata

hdfs namenode -bootstrapStandby

hadoop- daemon.sh start namenode

13. run the command in all node to start zookeeper

zkServer.sh start

14. run the command in all node to start datanodes

hadoop-daemon.sh start datanode

15 Format zookeeper failour controller in active namenode only

hdfs zkfc –formatZK

hadoop-daemon.sh start zkfc (in both)

To check namenode status from command line run the below command

hdfs haadmin –getServiceState namenode