Fully Distributed Cluster

sudo apt-get update

----------\*\*upload .pem key\*\*----------

windows : (open cmd)

pscp -i a.ppk a.pem ubuntu@52.18.223.48:/home/ubuntu/.ssh

linux : (terminal)

scp -i a.pem a.pem ubuntu@52.18.223.48:/home/ubuntu/.ssh

chmod 400 /home/ubuntu/.ssh/a.pem

-------\*\*configuring .profile\*\*-----------(make sure u are on nn)

nano ~/.profile

eval `ssh-agent` ssh-add /home/ubuntu/.ssh/a.pem

source ~/.profile

-------\*\*adding hosts\*\*---- (namenode , job tracker, secondary name node and 3 data nodes)

sudo nano /etc/hosts

172.31.23.8 ip-172-31-23-8.eu-west-1.compute.internal nn

172.31.23.7 ip-172-31-23-7.eu-west-1.compute.internal snn

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn1

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal dn2

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn3

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal jt

ssh snn

sudo nano /etc/hosts

172.31.23.8 ip-172-31-23-8.eu-west-1.compute.internal nn

172.31.23.7 ip-172-31-23-7.eu-west-1.compute.internal snn

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn1

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal dn2

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn3

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal jt

exit

ssh dn1

sudo nano /etc/hosts

172.31.23.8 ip-172-31-23-8.eu-west-1.compute.internal nn

172.31.23.7 ip-172-31-23-7.eu-west-1.compute.internal snn

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn1

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal dn2

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn3

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal jt

exit

ssh dn2

sudo nano /etc/hosts

172.31.23.8 ip-172-31-23-8.eu-west-1.compute.internal nn

172.31.23.7 ip-172-31-23-7.eu-west-1.compute.internal snn

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn1

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal dn2

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn3

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal jt

exit

ssh jt

sudo nano /etc/hosts

172.31.23.8 ip-172-31-23-8.eu-west-1.compute.internal nn

172.31.23.7 ip-172-31-23-7.eu-west-1.compute.internal snn

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn1

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal dn2

172.31.23.10 ip-172-31-23-10.eu-west-1.compute.internal dn13

172.31.23.9 ip-172-31-23-9.eu-west-1.compute.internal jt

exit

--------\*\*installing and configuring dsh\*\*--------------(make sure u are on nn)

ssh nn

sudo apt-get install dsh

cd /etc/dsh

sudo nano machines.list

#localhost

nn

jt

snn

dn1

dn2

dn2

---------------------------\*\*configuring bashrc linux env setup\*\*---------------(make sure u are on nn)

nano ~/.bashrc

export HADOOP\_PREFIX=/usr/local/hadoop/

export PATH=$PATH:$HADOOP\_PREFIX/bin

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export PATH=$PATH:$JAVA\_HOME

-----------------\*\*scp .pem .profile & .bashrc to all hosts\*\*------------------(make sure u are on nn)

scp .bashrc .profile ubuntu@snn:/home/ubuntu

scp .bashrc .profile ubuntu@jt:/home/ubuntu

scp .bashrc .profile ubuntu@dn1:/home/ubuntu

scp .bashrc .profile ubuntu@dn2:/home/ubuntu

scp .bashrc .profile ubuntu@dn3:/home/ubuntu

scp .ssh/a.pem ubuntu@snn:/home/ubuntu/.ssh

scp .ssh/abc.pem ubuntu@jt:/home/ubuntu/.ssh

scp .ssh/a.pem ubuntu@dn1:/home/ubuntu/.ssh

scp .ssh/a.pem ubuntu@dn2:/home/ubuntu/.ssh

scp .ssh/a.pem ubuntu@dn3:/home/ubuntu/.ssh

--------------\*\*ececute bash and .profile\*\*------------------------

dsh -a bash

dsh -a source ~/.profile

--------------\*\*changing host name\*\*-----------------------------\

sudo hostname nn

bash

ssh snn

sudo hostname snn

bash

ssh jt

sudo hostname jt

bash

ssh dn1

sudo hostname dn1

bash

ssh dn2

sudo hostname dn2

bash

ssh dn3

sudo hostname dn3

bash

ssh nn

-----------------------------\*\*downloading hadoop\*\*------------------------(make sure u are on nn)

dsh -a wget https://archive.apache.org/dist/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz

dsh -a tar -xzvf hadoop-1.2.1.tar.gz

dsh -a sudo mv hadoop-1.2.1 /usr/local/hadoop

----------\*\*downloading and installing java (Hadoop requires a working Java 1.6+)\*\*--------(make sure u are on nn)

dsh -a sudo apt-get update

dsh -a sudo apt-get install openjdk-8-jdk -y

------\*\*setting hadoop env\*\*------(make sure u are on nn)

nano /usr/local/hadoop/conf/hadoop-env.sh

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP\_OPTS=-Djava.net.preferIPV4Stack=true

-------\*\*configuring xml's\*\*------ (make sure u are on nn)

nano /usr/local/hadoop/conf/core-site.xml

<property>

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

<value>hdfs://nn:9000</value>

</property>

<property>

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

<value>/usr/local/hadoop/dfs</value>

</property>

⁃ nano /usr/local/hadoop/conf/hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

⁃ nano /usr/local/hadoop/conf/mapred-site.xml

<property>

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

<value>hdfs://jt:9001</value>

</property>

---------------\*\*configuring masters and slaves\*\*-------------(make sure u are on nn)

ssh nn

nano /usr/local/hadoop/conf/masters

#localhost

snn

ssh nn

nano /usr/local/hadoop/conf/slaves

#localhost

dn1

dn2

dn3

--------------\*\* scp xml’s to all hosts\*\*----------(make sure u are on nn)

cd /usr/local/hadoop/conf/

scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml slaves ubuntu@snn:/usr/local/hadoop/conf/

scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml slaves ubuntu@jt:/usr/local/hadoop/conf/

scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml slaves ubuntu@3dn:/usr/local/hadoop/conf/

scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml slaves ubuntu@1dn:/usr/local/hadoop/conf/

scp hadoop-env.sh core-site.xml hdfs-site.xml mapred-site.xml slaves ubuntu@2dn:/usr/local/hadoop/conf/

----------\*\*making tmp dir\*\*---------(make sure u are on nn)

ssh nn

dsh -a mkdir /usr/local/hadoop/node

-------------\*\*exec bash\*\*-------(make sure u are on nn)

dsh -a exec bash

----------------\*\*formatting hadoop namenode\*\*--------------(make sure u are on nn)

hadoop namenode -format

---------------\*\*starting logical (mapreduce ) demons\*\*------------------ (make sure u are on jt)

ssh jt

start-mapred.sh

----------\*\*starting dfs\*\*------------------- (make sure u are on nn)

ssh nn

start-dfs.sh

-------------------------\*\*check java process\*\*-----------

dsh -a jps

hadoop jar $HADOOP\_PREFIX/hadoop-\*examples\*.jar teragen 524288 input

hadoop jar $HADOOP\_PREFIX/hadoop-\*examples\*.jar terasort input output

hadoop jar $HADOOP\_PREFIX/hadoop-\*examples\*.jar teravalidate /user/ubuntu/output /user/ubuntu/validate

hadoop jar $HADOOP\_PREFIX/hadoop-\*test\*.jar TestDFSIO -write -nrFiles 5 -fileSize 50  
  
hadoop jar $HADOOP\_PREFIX/hadoop-\*test\*.jar TestDFSIO -read -nrFiles 5 -fileSize 50  
  
hadoop jar $HADOOP\_PREFIX/hadoop-\*test\*.jar TestDFSIO -clean

hadoop jar $HADOOP\_PREFIX/hadoop-\*test\*.jar mrbench -numRuns 50  
  
hadoop jar $HADOOP\_PREFIX/hadoop-\*examples\*.jar pi 10 100000