|  |
| --- |
| Checklist to Setup Hadoop Cluster |
|  |

|  |
| --- |
|  |



© Brain4ce Education Solutions Pvt. Ltd.

# Checklist to Setup Hadoop Cluster

Setup 3-4 linux machines in your vmbox. Each machine should have unique ip address and update in /etc/hosts without fail.

test ping across machines

java installed one each machine--java-1.6.0-openjdk-devel(if not installed ,use this command)

java -version: <- (does it show java-1.7)

rpm -qa | grep java

shows the java-1.7 pacakge

rpm -e package name

------------------------------------------------------------

openssh-server installed on each machine--yum install openssh-server

Downloads hadoop package

--->

--create a group and user across the machines

say hadoop--group

say hduser--userid

-->

generate ssh-keys

distribute across machines

check ssh access on machines

update .bashrc with java path

test$JAVA\_HOME

-----------------------

Downloads hadoop

-------------------------

Untar it in your location

Create a link to your untarred dir

then take your hadoop path and update in bashrc

change path to your hadoop path

:start editing config files....

Remember to make changes in etc/hosts as per

pseudo distributed

or fully distribute

---------------------------

edit

hadoop-env

core

mapred

hdfs

masters

slaves.....

bin/hadoop namenode -format

start your daemons:

bin/start-all.sh

bin/start-dfs.sh

bin/start-mapred.sh

or

bin/hadoop-daemon.sh start daemon name

merging of edits with fsimage

1. Restarting namenode

2. When snn contacts nn and does merging of edits into fsimage and pushing new fsimage back to nn

3. Using admin commands like saveNamespace ( which needs your namenode to be in safemode)

4. Manually telling snn to contact nn and do the merging---checkpoint force

Some important commands

bin/hadoop dfsadmin -report

bin/hadoop fs ( -ls -mkdir -copyFromLocal -put -D dfs.replication=value -rmr -lsr -cp)

bin/hadoop fsck / -blocks

bin/hadoop fsck / -blocks -files

bin/hadoop dfsadmin -saveNamespace

bin/hadoop dfsadmin -safemode ( enter get leave wait)

simulation of under replicated blocks

writing file with different replication -D dfs.replication=value

bin/hadoop dfsadmin -metasave

bin/hadoop secondarynamenode -checkpoint force

====================================================