**Assignment 1:**

**Problem Statement:**

**Create a Hadoop cluster in local mode on centos**

1)Install centos 6.8 in VMware workstation from ISO image.

Okay, so we are going to set up a Single Node Cluster today.

2) We will have 1 machines. It will run Namenode, one Secondary Namenode,one Job Tracker,Datanodes and Tasktrackers.

3) disable ipv6 and selinux

Edit /boot/grub/grub.conf and append below parameters in the end of the kernel line of default kernel & reboot the system.

ipv6.disable=1 selinux=0

4) disable

chkconfig iptables off

chkconfig ip6tables off

chkconfig NetworkManager off

chkconfig network on

5) download oracle jdk-1.6.x from below url

<http://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5/RPMS/x86_64/jdk-6u31-linux-amd64.rpm>

6) download hadoop-1.2.1 from below url

<https://archive.apache.org/dist/hadoop/core/hadoop-1.2.1/hadoop-1.2.1.tar.gz>

7) install jdk & set alternatives

Yum localinstall jdk-6u31-linux-amd64.rpm -y

alternatives --install /usr/bin/java java /usr/java/jdk1.6.0\_31/bin/java 210000

java -version

8) create a local group and user for hadoop installation.

groupadd -g 1001 hadoop

useradd -u 1001 -d /usr/local/hadoop -s /bin/bash -g hadoop hduser

passwd hduser

cp /etc/skel/.bash\* /usr/local/hadoop

chown -R hduser:hadoop /usr/local/hadoop/.bash\*

9) extract hadoop tar and copy content to /usr/local/hadoop

tar xzf hadoop-1.2.1.tar.gz

cp hadoop-1.2.1/\* /usr/local/hadoop

chown -R hduser:hadoop /usr/local/hadoop/\*

10) set .bashrc of hduser (add below content after last line of the file)

#su - hduser

$vi .bashrc

export JAVA\_HOME=/usr/java/jdk1.6.0\_31

export HADOOP\_HOME=/usr/local/hadoop

export PATH=$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$JAVA\_HOME/bin:$PATH

export HADOOP\_INSTALL=/usr/local/hadoop

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

export PATH=$PATH:$HADOOP\_INSTALL/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

export YARN\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

11) setup passwordless env for hduser

#su - hduser

$ssh-keygen

$ssh-copy-id -i ~/.ssh/id\_rsa.pub hduser@localhost

12) Edit ssh\_config file and add below section in the end of the file

#vi /etc/ssh/ssh\_config

StrictHostKeyChecking no

13) Edit hadoop-env.sh

Enter the java home

14) Now start configuring hadoop config file as below in "/usr/local/hadoop/conf" dir

core-site.xml

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

[root@sandbox1 ~]# cat /usr/local/hadoop/conf/core-site.xml

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

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

<value>/data/tmp</value>

</property>

<property>

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

<value>hdfs://localhost:8020</value>

</property>

</configuration>

[root@sandbox1 ~]#

hdfs-site.xml

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

[root@sandbox1 ~]# cat /usr/local/hadoop/conf/hdfs-site.xml

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>file:/data/nn</value>

</property>

<property>

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

<value>file:/data/dn1,file:/data/dn2,file:/data/dn3</value>

</property>

<property>

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

<value>file:/data/snn</value>

</property>

<property>

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

<value>file:/data/snn</value>

</property>

</configuration>

mapred-site.xml

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

[root@sandbox1 ~]# cat /usr/local/hadoop/conf/mapred-site.xml

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

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

<value>localhost:8021</value>

</property>

</configuration>

15) created required dirs and give correct file permission.

#mkdir /data/{tmp,nn,dn1,dn2,dn3,snn}

#chown -R hduser:hadoop /data

16) now format namenode

#su - hduser

$hadoop namenode –format

17) start hdfs

$ cd /usr/local/hadoop/bin

$./start-dfs.sh

18) open web browser and check hdfs file system

<http://localhost:50070>

19) start mapreduce daemon

$cd /usr/local/hadoop/bin

$./start-mapred.sh

20) check mapreduce service via web browser

<http://localhost:50030>

21) create required hdfs user dir to run MR job

#su - hduser

$hadoop fs -mkdir /user/hduser

$hadoop fs -mkdir /tmp

22) test sample mr job

$hadoop jar /usr/local/hadoop/hadoop-examples-1.2.1.jar pi 1 1

**Client configuration will be done in our projects as suggested by our mentor.**