Apache hadoop cluster multinode

***Vikram******Iyer***

*12-10-2014*

**Hosts file**

**vi /etc/hosts**

10.77.36.1 hmaster

10.77.36.1 hslave1

10.77.36.1 hslave2

**Create user and group**

**root@hmaster:~$ useradd user**

**root@hmaster:~$ groupadd hadoop**

**root@hmaster:~$ usermod -a -G hadoop user**

**root@hmaster:~$ usermod -a -G hadoop root**

**Passwordless ssh**

**user@hmaster:~$ ssh-keygen**

**user@hmaster:~$ ssh-copy-id -i ~/.ssh/id\_rsa.pub user@hslave1**

**user@hmaster:~$ ssh-copy-id -i ~/.ssh/id\_rsa.pub user@hslave2**

**Download tar and create hadoop directories**

**root@hmaster:~$ mkdir /opt/user**

**root@hmaster:~$ wget --no-cookies --no-check-certificate --header "Cookie: gpw\_e24=http%3A%2F%2Fwww.oracle.com%2F; oraclelicense=accept-securebackup-cookie" "**[**http://download.oracle.com/otn-pub/java/jdk/7u79-b15/jdk-7u79-linux-x64.tar.gz**](http://download.oracle.com/otn-pub/java/jdk/7u79-b15/jdk-7u79-linux-x64.tar.gz)**" -O /opt/user/downloads/jdk-7u79-linux-x64.tar.gz**

**root@hmaster:~$ wget** <http://apache.mirrors.pair.com/hadoop/common/hadoop-2.5.2/hadoop-2.5.2.tar.g>

**root@hmaster: /opt/user $ tar -xzf hadoop-2.5.2.tar.gz**

**root@hmaster: /opt/user $ tar -xzf jdk-7u71-linux-x64.tar.gz**

**root@hmaster: /opt/user $ mv hadoop-2.5.2 hadoop**

**root@hmaster: /opt/user $ mkdir -p /opt/user/hadoop/hdfs/tmp**

**root@hmaster: /opt/user $ chown -R user:hadoop /opt/user**

**root@hmaster: /opt/user $ chmod -R 755 /opt/user/hadoop**

**Configurations**

(conf-dir = $hadoop\_home/etc/hadoop, hadoop\_home = /opt/user/hadoop)

**Note**: (mandatory on all machines in cluster)

1. slaves

2. core-site.xml

3. hdfs-site.xml

4. yarn-site.xml

5. mapred-site.xml

**Slaves**

This file will decide all the datanodes that will be running in the cluster

hslave1

hslave2

**core-site.xml**

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://hmaster:9000</value>

</property>

<property>

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

<value>/opt/user/hadoop/hdfs/tmp</value>

</property>

</configuration>

**hdfs-site.xml**

<configuration>

<property>

<name>dfs.permissions.superusergroup</name>

<value>hadoop</value>

</property>

<property>

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

<value>file:/opt/user/hadoop/yarn/yarn\_data/hdfs/namenode</value>

</property>

<property>

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

<value>file:/opt/user/hadoop/yarn/yarn\_data/hdfs/datanode</value>

</property>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

**yarn-site.xml**

<configuration>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>hmaster</value>

</property>

<property><name>yarn.resourcemanager.address</name>

<value>hmaster</value>

</property>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>hmaster</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

**mapred-site.xml**

<configuration>

<property>

<name>mapreduce.jobhistory.address</name>

<value>hmaster:10020</value>

</property>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

**bashrc**

**Note**: (mandatory on all machines in cluster)

**root@machine $ vi /root/.bashrc**

export HADOOP\_HOME=/opt/user/hadoop/

export JAVA\_HOME=/opt/user/jdk1.7.0\_71/

export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop

export HADOOP\_YARN\_HOME=$HADOOP\_HOME

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

**root@machine $ vi /home/shareinsights/.bashrc**

export HADOOP\_HOME=/opt/user/hadoop/

export JAVA\_HOME=/opt/user/jdk1.7.0\_71/

export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop

export HADOOP\_YARN\_HOME=$HADOOP\_HOME

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

**root@machine $ source /root/.bashrc**

**root@machine $ source /home/user/.bashrc**

**Format the namenode and start the daemons**

**root@hmaster $ su – user ; hadoop namenode -format**

**user@hmaster $ jps**

22611 Jps

**user@hmaster $ start-dfs.sh**

**user@hmaster $ start-yarn.sh**

**user@hmaster $ mr-jobhistory-daemon.sh start historyserver**

**user@hmaster $ jps**

17289 NodeManager

22611 Jps

16982 SecondaryNameNode

17159 ResourceManager

17689 JobHistoryServer

16624 NameNode

16787 DataNode

**user@hmaster $ hadoop jar /opt/user/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.2.jar pi 2 4**

(run and check if the sample hadoop example is running)

Your hadoop cluster is ready to be used.