How to Set Up Hadoop Multi-Node Cluster on CentOS 7/6

By Rahul — July 22, 2013 4 Mins Read

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models.

Our earlier article about hadoop was describing to how to setup single node cluster. This article will help you for step by step installing and configuring Hadoop Multi-Node Cluster on CentOS/RHEL 6.

Setup Details:

Hadoop Master: 192.168.1.15 (hadoop-master)
Hadoop Slave : 192.168.1.16 (hadoop-slave-1)
Hadoop Slave : 192.168.1.17 (hadoop-slave-2)

Step 1. Install Java

Before installing hadoop make sure you have java installed on all nodes of hadoop cluster systems.

```
# java -version

java version "1.7.0_75"
Java(TM) SE Runtime Environment (build 1.7.0_75-b13)
Java HotSpot(TM) 64-Bit Server VM (build 24.75-b04, mixed mode)
```

If you do not have java installed use following article to install Java.

Steps to install JAVA 7 on CentOS/RHEL 7/6/5

Step 2. Create User Account

Create a system user account on both master and slave systems to use for hadoop installation

```
# useradd hadoop
# passwd hadoop
```

```
Changing password for user hadoop.
New password:
Retype new password:
passwd: all authentication tokens updated successfully.
```

Step 3: Add FQDN Mapping

Edit **/etc/hosts** file on all master and slave servers and add following entries.

```
# vim /etc/hosts
```

```
192.168.1.15 hadoop-master
192.168.1.16 hadoop-slave-1
192.168.1.17 hadoop-slave-2
```

Step 4. Configuring Key Based Login

It's required to set up hadoop user to ssh itself without password. Use following commands to configure auto login between all hadoop cluster servers..

```
# su - hadoop
$ ssh-keygen -t rsa
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop@hadoop-master
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop@hadoop-slave-1
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop@hadoop-slave-2
$ chmod 0600 ~/.ssh/authorized_keys
$ exit
```

Step 5. Download and Extract Hadoop Source

Download hadoop latest available version from its official site at hadoop-master server only.

```
# mkdir /opt/hadoop
# cd /opt/hadoop/
# wget http://apache.mesi.com.ar/hadoop/common/hadoop-1.2.0/hadoop-1.2.0.tar.gz
# tar -xzf hadoop-1.2.0.tar.gz
# mv hadoop-1.2.0 hadoop
# chown -R hadoop /opt/hadoop
# cd /opt/hadoop/hadoop/
```

Step 6: Configure Hadoop

First edit hadoop configuration files and make following changes.

6.1 Edit core-site.xml

```
# vim conf/core-site.xml
```

```
#Add the following inside the configuration tag
<property>
    <name>fs.default.name</name>
    <value>hdfs://hadoop-master:9000/</value>
</property>
<property>
    <name>dfs.permissions</name>
    <value>false</value>
</property>
```

6.2 Edit hdfs-site.xml

```
# vim conf/hdfs-site.xml
```

```
# Add the following inside the configuration tag
<property>
        <name>dfs.data.dir</name>
        <value>/opt/hadoop/hadoop/dfs/name/data</value>
        <final>true</final>
</property>
```

```
<property>
        <name>dfs.name.dir</name>
        <value>/opt/hadoop/hadoop/dfs/name</value>
        <final>true</final>
</property>
<property>
        <name>dfs.replication</name>
        <value>1</value>
</property>
```

6.3 Edit mapred-site.xml

```
# vim conf/mapred-site.xml
```

```
# Add the following inside the configuration tag
<property>
        <name>mapred.job.tracker</name>
        <value>hadoop-master:9001</value>
</property>
```

6.4 Edit hadoop-env.sh

```
# vim conf/hadoop-env.sh
```

```
export JAVA_HOME=/opt/jdk1.7.0_75
export HADOOP_OPTS=-Djava.net.preferIPv4Stack=true
export HADOOP_CONF_DIR=/opt/hadoop/hadoop/conf
```

Set JAVA_HOME path as per your system configuration for java.

Step 7: Copy Hadoop Source to Slave Servers

After updating above configuration, we need to copy the source files to all slaves servers.

```
# su - hadoop
$ cd /opt/hadoop
$ scp -r hadoop hadoop-slave-1:/opt/hadoop
$ scp -r hadoop hadoop-slave-2:/opt/hadoop
```

Step 8: Configure Hadoop on Master Server Only

Go to hadoop source folder on hadoop-master and do following settings.

```
# su - hadoop
$ cd /opt/hadoop/hadoop
```

```
$ vim conf/masters

hadoop-master
```

```
$ vim conf/slaves

hadoop-slave-1
hadoop-slave-2
```

Format Name Node on Hadoop Master only

```
# su - hadoop
$ cd /opt/hadoop/hadoop
$ bin/hadoop namenode -format
```

```
13/07/13 10:58:07 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = hadoop-master/192.168.1.15
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 1.2.0
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/bra
STARTUP_MSG:   java = 1.7.0_25
```

Step 9: Start Hadoop Services

Use the following command to start all hadoop services on Hadoop-Master

```
$ bin/start-all.sh
```