

Big Data - Concepts

22 December 2023 09:05

Implementation of Multi-Node Hadoop Cluster Contd...,

Hadoop Master-Slave Configuration Setup

1. Configure Hadoop in an master-slave setup

Perform the following activity on all the nodes

```
root@slave1:/home/hadoop
hadoop@slave1:~$ pwd
/home/hadoop
hadoop@slave1:~$ ls
hadoop-2.4.1.tar.gz jdk-7u80-linux-x64.tar.gz
hadoop@slave1:~$ su
Password:
root@slave1:/home/hadoop# tar -zxf hadoop-2.4.1.tar.gz
root@slave1:/home/hadoop# tar -zxf jdk-7u80-linux-x64.tar.gz
root@slave1:/home/hadoop# ls -l
total 285352
drwxr-xr-x 9 67974 users 4096 Jun 21 2014 hadoop-2.4.1
-rw-rw-r-- 1 hadoop hadoop 138656756 Dec 22 04:22 hadoop-2.4.1.tar.gz
drwxr-xr-x 8 uucp 143 4096 Apr 11 2015 jdk1.7.0_80
-rw-rw-r-- 1 hadoop hadoop 153530841 Dec 22 04:22 jdk-7u80-linux-x64.tar.gz
root@slave1:/home/hadoop# mv hadoop-2.4.1 /usr/local/hadoop
root@slave1:/home/hadoop# mv jdk1.7.0_80 /usr/local/
root@slave1:/home/hadoop# chown -R hadoop.hadoop /usr/local/hadoop
root@slave1:/home/hadoop# chown -R root.root /usr/local/jdk1.7.0_80

Command 'chonw' not found, did you mean:

  command 'chown' from deb coreutils (8.30-3ubuntu2)

Try: apt install <deb name>

root@slave1:/home/hadoop# chown -R root.root /usr/local/jdk1.7.0_80
root@slave1:/home/hadoop#
```

```

hadoop@slave2:~$ ls
hadoop-2.4.1.tar.gz jdk-7u80-linux-x64.tar.gz
hadoop@slave2:~$ tar -zxf jdk-7u80-linux-x64.tar.gz
hadoop@slave2:~$ tar -zxf hadoop-2.4.1.tar.gz
hadoop@slave2:~$ ls -l
total 285356
drwxr-xr-x 9 hadoop hadoop 4096 Jun 21 2014 hadoop-2.4.1
-rw-rw-r-- 1 hadoop hadoop 138656756 Dec 22 04:22 hadoop-2.4.1.tar.gz
drwxr-xr-x 8 hadoop hadoop 4096 Apr 11 2015 jdk1.7.0_80
-rw-rw-r-- 1 hadoop hadoop 153530841 Dec 22 04:22 jdk-7u80-linux-x64.tar.gz
hadoop@slave2:~$ su
Password:
root@slave2:/home/hadoop# mv hadoop-2.4.1 /usr/local/hadoop
root@slave2:/home/hadoop# mv jdk1.7.0_80 /usr/local/
root@slave2:/home/hadoop# chown -R hadoop.hadoop /usr/local/hadoop
root@slave2:/home/hadoop# chown -R root.root /usr/local/jdk1.7.0_80/
root@slave2:/home/hadoop# exit
exit
hadoop@slave2:~$ █

```

```

# nano /etc/bash.bashrc
<add the following lines to the end of the file >

export JAVA_HOME=/usr/local/jdk1.7.0_80
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_INSTALL=$HADOOP_HOME
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

```

<save and exit>

Logout and login as hadoop in both the nodes and run the following command

```
# java -version
# hadoop version
```

we should be able to get the following output after executing the above command on both the nodes

```

hadoop@slave1:~$ 
hadoop@slave1:~$ java -version
java version "1.7.0_80"
Java(TM) SE Runtime Environment (build 1.7.0_80-b15)
Java HotSpot(TM) 64-Bit Server VM (build 24.80-b11, mixed mode)
hadoop@slave1:~$ hadoop version
Hadoop 2.4.1
Subversion http://svn.apache.org/repos/asf/hadoop/common -r 1604318
Compiled by jenkins on 2014-06-21T05:43Z
Compiled with protoc 2.5.0
From source with checksum bb7ac0a3c73dc131f4844b873c74b630
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-2.4.1.jar

```

```
hadoop@slave2:~$ java -verion
Unrecognized option: -verion
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
hadoop@slave2:~$ hadoop version
Hadoop 2.4.1
Subversion http://svn.apache.org/repos/asf/hadoop/common -r 1604318
Compiled by jenkins on 2014-06-21T05:43Z
Compiled with protoc 2.5.0
From source with checksum bb7ac0a3c73dc131f4844b873c74b630
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-2.4.1.jar
hadoop@slave2:~$ echo $JAVA_HOME
/usr/local/jdk1.7.0_80
hadoop@slave2:~$ echo $HADOOP_HOME
/usr/local/hadoop
hadoop@slave2:~$
```

Perform the activity on all the nodes

```
$ su
# mkdir -p /opt/hadoop
# chown -R hadoop.hadoop /opt/hadoop
# exit
```

On MainServer i.e, masternode

```
root@mainserver1:/usr/local/hadoop/etc/hadoop#
root@mainserver1:/opt# cd $HADOOP_HOME/etc/hadoop
root@mainserver1:/usr/local/hadoop/etc/hadoop# nano core-site.xml
root@mainserver1:/usr/local/hadoop/etc/hadoop# tail core-site.xml
<configuration>
    <property>
        <name>fs.default.name</name>
        <value>hdfs://mainserver1:9000</value>
    </property>
    <property>
        <name>dfs.permissions</name>
        <value>false</value>
    </property>
</configuration>
root@mainserver1:/usr/local/hadoop#
```

```
root@mainserver1:/usr/local/hadoop/etc/hadoop# nano hdfs-site.xml
root@mainserver1:/usr/local/hadoop/etc/hadoop# tail -20 hdfs-site.xml
-->

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

<configuration>
<property>
    <name>dfs.replication</name>
    <value>1</value>
</property>

<property>
    <name>dfs.name.dir</name>
    <value>/opt/hadoop/hadoopinfra/hdfs/namenode </value>
</property>

<property>
    <name>dfs.data.dir</name>
    <value>/opt/hadoop/hadoopinfra/hdfs/datanode </value>
</property>
</configuration>
```

```

root@mainserver1:/usr/local/hadoop/etc/hadoop# nano mapred-site.xml
root@mainserver1:/usr/local/hadoop/etc/hadoop# tail mapred-site.xml
-->

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

<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>mainserver1:9001</value>
  </property>
</configuration>
root@mainserver1:/usr/local/hadoop# 

```

```

hadoop@mainserver1:/usr/local/hadoop/etc/hadoop$ nano yarn-site.xml
hadoop@mainserver1:/usr/local/hadoop/etc/hadoop$ tail yarn-site.xml
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<!-- Site specific YARN configuration properties -->

  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
hadoop@mainserver1:/usr/local/hadoop/etc/hadoop$ 

```

Edit hadoop-env.sh and add the following entry - on all the nodes

```

# cd $HADOOP_HOME/etc/hadoop
# nano hadoop-env.sh
# The java implementation to use.
export JAVA_HOME=/usr/local/jdk1.7.0_80

```

```

root@mainserver1:/usr/local/hadoop/etc/hadoop# ls
capacity-scheduler.xml      hdfs-site.xml
configuration.xsl           httpfs-env.sh
container-executor.cfg       httpfs-log4j.properties
core-site.xml                httpfs-signature.secret
hadoop-env.cmd               httpfs-site.xml
hadoop-env.sh                log4j.properties
hadoop-metrics2.properties  mapred-env.cmd
hadoop-metrics.properties   mapred-env.sh
hadoop-policy.xml            mapred-queues.xml.template
root@mainserver1:/usr/local/hadoop/etc/hadoop# cat slaves
localhost
root@mainserver1:/usr/local/hadoop/etc/hadoop# nano masters
root@mainserver1:/usr/local/hadoop/etc/hadoop# cat masters
mainserver1
root@mainserver1:/usr/local/hadoop/etc/hadoop# nano slaves
root@mainserver1:/usr/local/hadoop/etc/hadoop# cat slaves
slave1
slave2
root@mainserver1:/usr/local/hadoop/etc/hadoop# cd $HADOOP_HOME/etc/hadoop
root@mainserver1:/usr/local/hadoop/etc/hadoop# 

```

Execute the following command only on mainnode ie., master node only

```

root@mainserver1:/usr/local/hadoop/bin
root@mainserver1:/usr/local/hadoop/etc/hadoop# cd $HADOOP_HOME/bin
root@mainserver1:/usr/local/hadoop/bin# hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

23/12/22 08:40:16 INFO namenode.NameNode: STARTUP_MSG:
/*****STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = mainserver1/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.4.1
STARTUP_MSG: classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/jets3t-0.9.0.jar:/usr/local/hadoop/share/hadoop/common/lib/zookeeper-3.4.5.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/usr/local/hadoop/share/hadoop/common/lib/jsch-0.1.42.jar:/usr/local/hadoop/share/hadoop/common/lib/xmlenc-0.52.jar:/usr/local/hadoop/share/hadoop/common/lib/activation-1.1.jar:/usr/local/hadoop/share/hadoop/common/lib/slf4j-api-1.7.5.jar:/usr/local/hadoop/share/hadoop/common/lib/jettison-1.1.jar:/usr/local/hadoop/share/hadoop/common/lib/guava-11.0.2.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-xc-1.8.8.jar:/usr/local/hadoop/share/hadoop/common/lib/jsr305-1.3.9.jar:/usr/local/hadoop/share/hadoop/common/lib/servlet-api-2.5.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-collections-3.2.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jersey-core-1.9.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-io-2.4.jar:/usr/local/hadoop/share/hadoop/common/lib/jetty-6.1.26.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-jaxrs-1.8.8.jar:/usr/local/hadoop/share/hadoop/common/lib/httpcore-4.2.5.jar:/usr/local/hadoop/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-mapper-asl-1.8.8.jar:/usr/local/hadoop/share/hadoop/common/lib/asm-3.2.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-compress-1.4.1.jar:/usr/local/hadoop/share/hadoop/common/lib/java-xmlbuilder-0.4.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-digester-1.8.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-lang-2.6.jar:/usr/local/hadoop/share/hadoop/common/lib/jsp-api-2.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jetty-util-6.1.26.jar:/usr/local/hadoop/share/hadoop/common/lib/mockito-all-1.8.5.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-math3-3.1.1.jar:/usr/local
root@mainserver1:/usr/local/hadoop/bin
23/12/22 08:40:21 INFO util.GSet: VM type      = 64-bit
23/12/22 08:40:21 INFO util.GSet: 1.0% max memory 966.7 MB = 9.7 MB
23/12/22 08:40:21 INFO util.GSet: capacity      = 2^20 = 1048576 entries
23/12/22 08:40:21 INFO namenode.NameNode: Caching file names occurring more than 10 times
23/12/22 08:40:21 INFO util.GSet: Computing capacity for map cachedBlocks
23/12/22 08:40:21 INFO util.GSet: VM type      = 64-bit
23/12/22 08:40:21 INFO util.GSet: 0.25% max memory 966.7 MB = 2.4 MB
23/12/22 08:40:21 INFO util.GSet: capacity      = 2^18 = 262144 entries
23/12/22 08:40:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
23/12/22 08:40:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
23/12/22 08:40:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension      = 30000
23/12/22 08:40:21 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
23/12/22 08:40:21 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
23/12/22 08:40:21 INFO util.GSet: Computing capacity for map NameNodeRetryCache
23/12/22 08:40:21 INFO util.GSet: VM type      = 64-bit
23/12/22 08:40:21 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 297.0 KB
23/12/22 08:40:21 INFO util.GSet: capacity      = 2^15 = 32768 entries
23/12/22 08:40:21 INFO namenode.AclConfigFlag: ACLs enabled? false
23/12/22 08:40:22 INFO namenode.FSImage: Allocated new BlockPoolId: BP-424479396-127.0.1.1-1703234421708
23/12/22 08:40:22 INFO common.Storage: Storage directory /opt/hadoop/hadoopinfra/hdfs/namenode has been successfully formatted.
23/12/22 08:40:22 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
23/12/22 08:40:22 INFO util.ExitUtil: Exiting with status 0
23/12/22 08:40:22 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****SHUTDOWN_MSG: Shutting down NameNode at mainserver1/127.0.1.1
*****/
root@mainserver1:/usr/local/hadoop/bin#

```

Note: We can also run the following command \$ hdfs namenode -format

\$ start-all.sh {execute the command as an hadoop user }

```

root@mainserver1:/usr/local/hadoop/bin# exit
exit
hadoop@mainserver1:/opt$ cd $HADOOP_HOME/bin
hadoop@mainserver1:/usr/local/hadoop/bin$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
23/12/22 09:27:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [mainserver1]
The authenticity of host 'mainserver1 (127.0.1.1)' can't be established.
ECDSA key fingerprint is SHA256:AojXpepX/dNKoHN/t6wzgvHeXgCYE2SOJ8mq0O9III1s.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
mainserver1: Warning: Permanently added 'mainserver1' (ECDSA) to the list of known hosts.
mainserver1: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hadoop-namenode-mainserver1.out
slave2: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hadoop-datanode-slave2.out
slave1: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hadoop-datanode-slave1.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hadoop-secondarynamenode-mainserver1.out
23/12/22 09:28:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hadoop-resourcemanager-mainserver1.out
slave1: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hadoop-nodemanager-slave1.out
slave2: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hadoop-nodemanager-slave2.out
hadoop@mainserver1:/usr/local/hadoop/bin$ 

```

```

192.168.56.100 - PUTTY
hadoop@mainserver1:/usr/local/hadoop/bin$ jps
2017 Jps
1744 ResourceManager
1616 SecondaryNameNode
hadoop@mainserver1:/usr/local/hadoop/bin$ ss -ant
State      Recv-Q      Send-Q      Local Address:Port          Peer Address:Port      Process
LISTEN      0           128          0.0.0.0:50090          0.0.0.0:*
LISTEN      0           4096         127.0.0.53:lo:53        0.0.0.0:*
LISTEN      0           128          0.0.0.0:22           0.0.0.0:*
ESTAB       0           0            192.168.56.100:39512    192.168.56.102:22
ESTAB       0           0            192.168.56.100:39294    192.168.56.101:22
ESTAB       0           0            192.168.56.100:22        192.168.56.1:50140
ESTAB       0           64           192.168.56.100:22        192.168.56.1:50073
ESTAB       0           0            192.168.56.100:22        192.168.56.1:50096
LISTEN      0           128           *:8032                  *:*
LISTEN      0           128           *:8033                  *:*
LISTEN      0           128           [::]:22                 [::]:*
LISTEN      0           128           *:8088                  *:*
LISTEN      0           128           *:8030                  *:*
LISTEN      0           128           *:8031                  *:*

```

```

hadoop@mainserver1:~$ cd /opt
hadoop@mainserver1:/opt$ tree
.
└── hadoop
    └── hadoopinfra
        └── hdfs
            └── namenode
                └── current
                    ├── fsimage_00000000000000000000
                    ├── fsimage_00000000000000000000.md5
                    ├── seen_txid
                    └── VERSION

5 directories, 4 files
hadoop@mainserver1:/opt$ 

```

```

hadoop@slave1:~$ jps
1489 Jps
1329 NodeManager
1167 DataNode
hadoop@slave1:~$ ss -ant
State Recv-Q Send-Q Local Address:Port          Peer Address:Port    Process
LISTEN  0      4096   127.0.0.53%lo:53          0.0.0.0:*
LISTEN  0      128    0.0.0.0:22              0.0.0.0:*
LISTEN  0      50    0.0.0.0:50010            0.0.0.0:*
LISTEN  0      128    0.0.0.0:50075            0.0.0.0:*
LISTEN  0      128    0.0.0.0:50020            0.0.0.0:*
ESTAB  0      0      192.168.56.101:22        192.168.56.100:39294
LISTEN  0      128    *:8042                  *:*
LISTEN  0      128    [::]:22                 [::]:*
LISTEN  0      50    *:13562                *:*
LISTEN  0      128    *:43111                *:*
LISTEN  0      128    *:8040                 *:*
hadoop@slave1:~$ cd /opt
hadoop@slave1:/opt$ tree
.
└── hadoop
    └── hadoopinfra
        └── hdfs
            └── datanode

4 directories, 0 files
hadoop@slave1:/opt$ 

```

```

root@slave2:~$ jps
1169 DataNode
1468 Jps
1335 NodeManager
hadoop@slave2:~$ ss -ant
State Recv-Q Send-Q Local Address:Port          Peer Address:Port    Process
LISTEN  0      4096   127.0.0.53%lo:53          0.0.0.0:*
LISTEN  0      128    0.0.0.0:22              0.0.0.0:*
LISTEN  0      50    0.0.0.0:50010            0.0.0.0:*
LISTEN  0      128    0.0.0.0:50075            0.0.0.0:*
LISTEN  0      128    0.0.0.0:50020            0.0.0.0:*
ESTAB  0      0      192.168.56.102:22        192.168.56.100:39512
LISTEN  0      128    [::]:22                 [::]:*
LISTEN  0      50    *:13562                *:*
LISTEN  0      128    *:8040                 *:*
LISTEN  0      128    *:8042                 *:*
LISTEN  0      128    *:34413                *:*

```

```

root@slave2:~# cd /oot
-bash: cd: /oot: No such file or directory
root@slave2:~# cd /opt
root@slave2:/opt# tree
.
└── hadoop
    └── hadoopinfra
        └── hdfs
            └── datanode

4 directories, 0 files
root@slave2:/opt# 

```