Big Data - Concepts

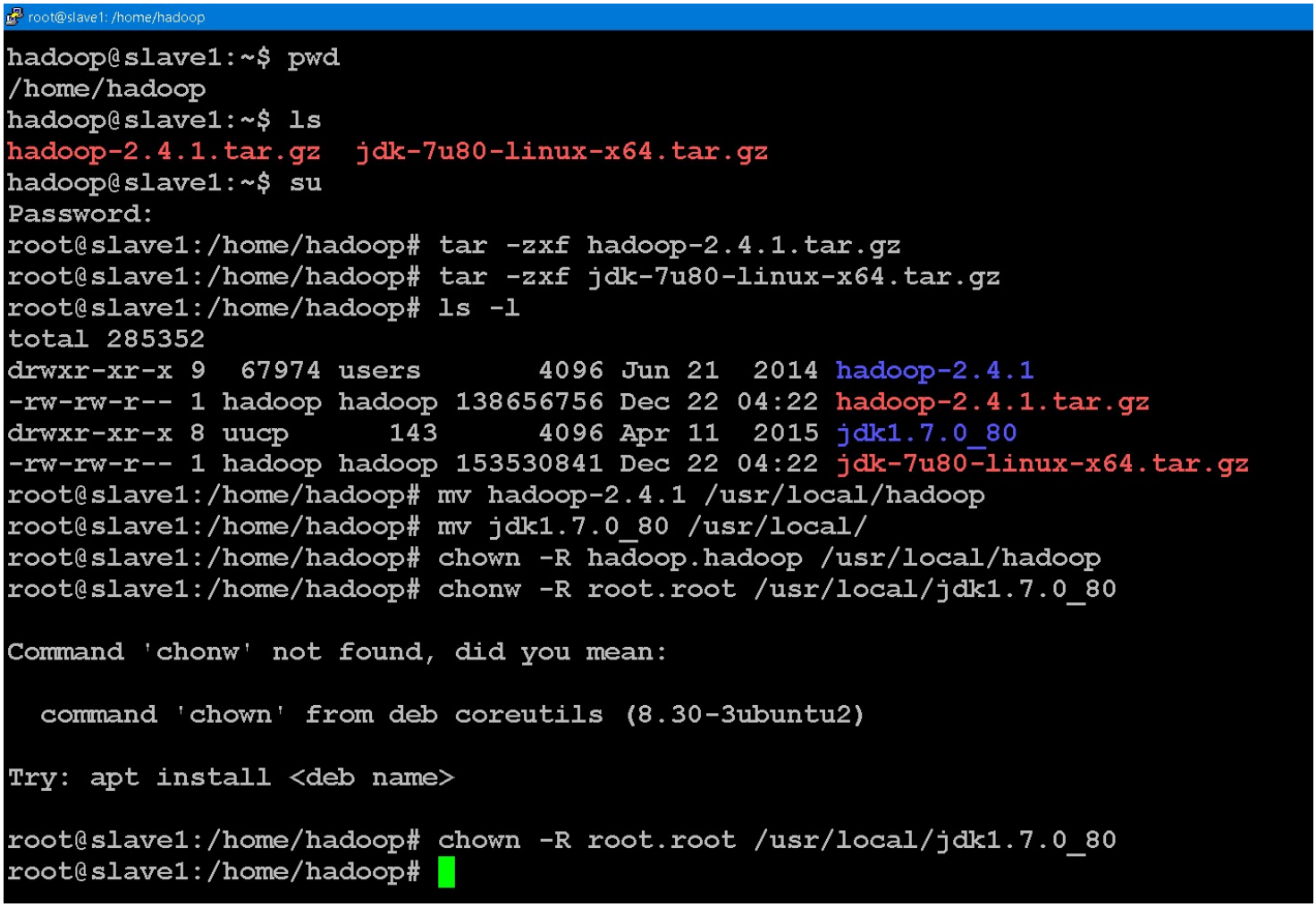
22 December 2023 09:05

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



```
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: ~
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/etc/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/etc/hadoop#
```

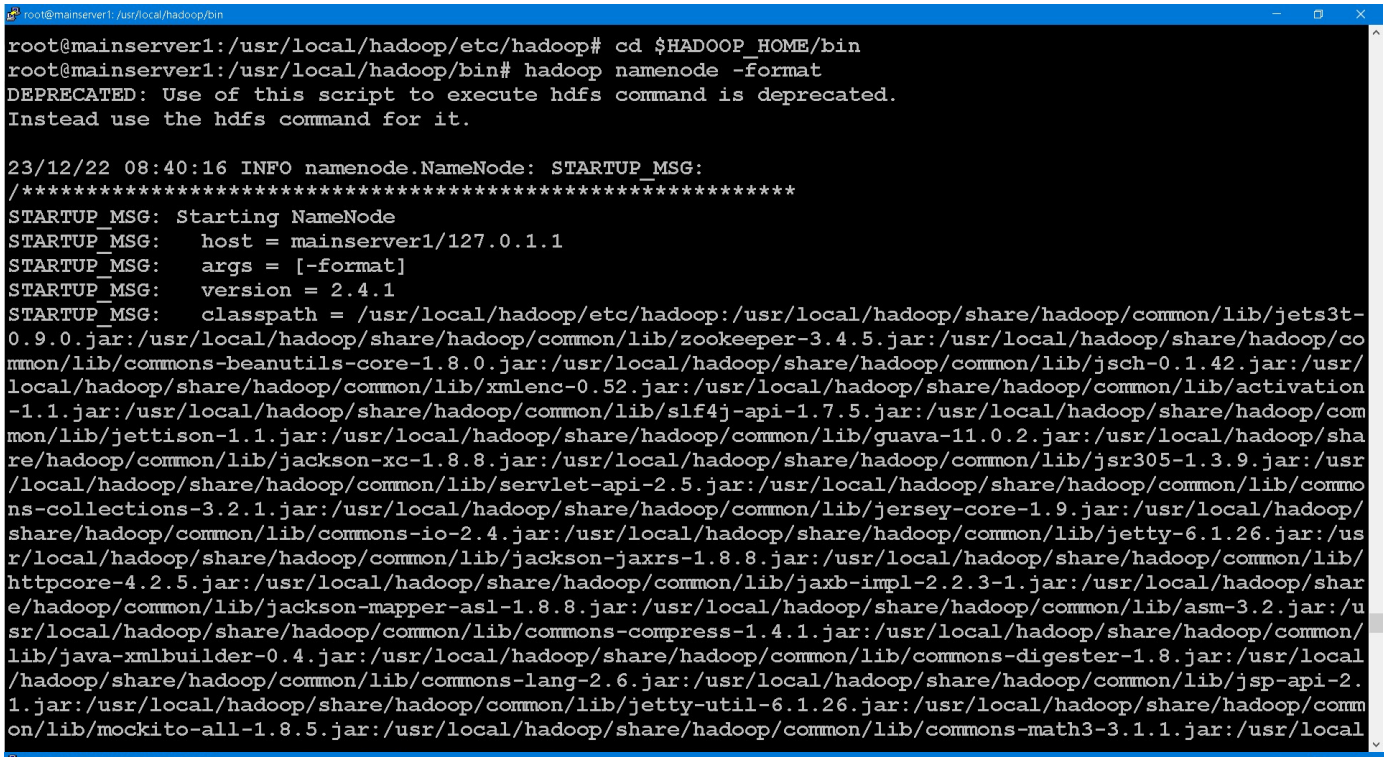
```
hadoop@mainserver1:/usr/local/hadoop/etc/hadoop$ nano yarn-site.xml
hadoop@mainserver1:/usr/local/hadoop/etc/hadoop$ tail  yarn-site.xml
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<!-- Site specific YARN configuration properties -->

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

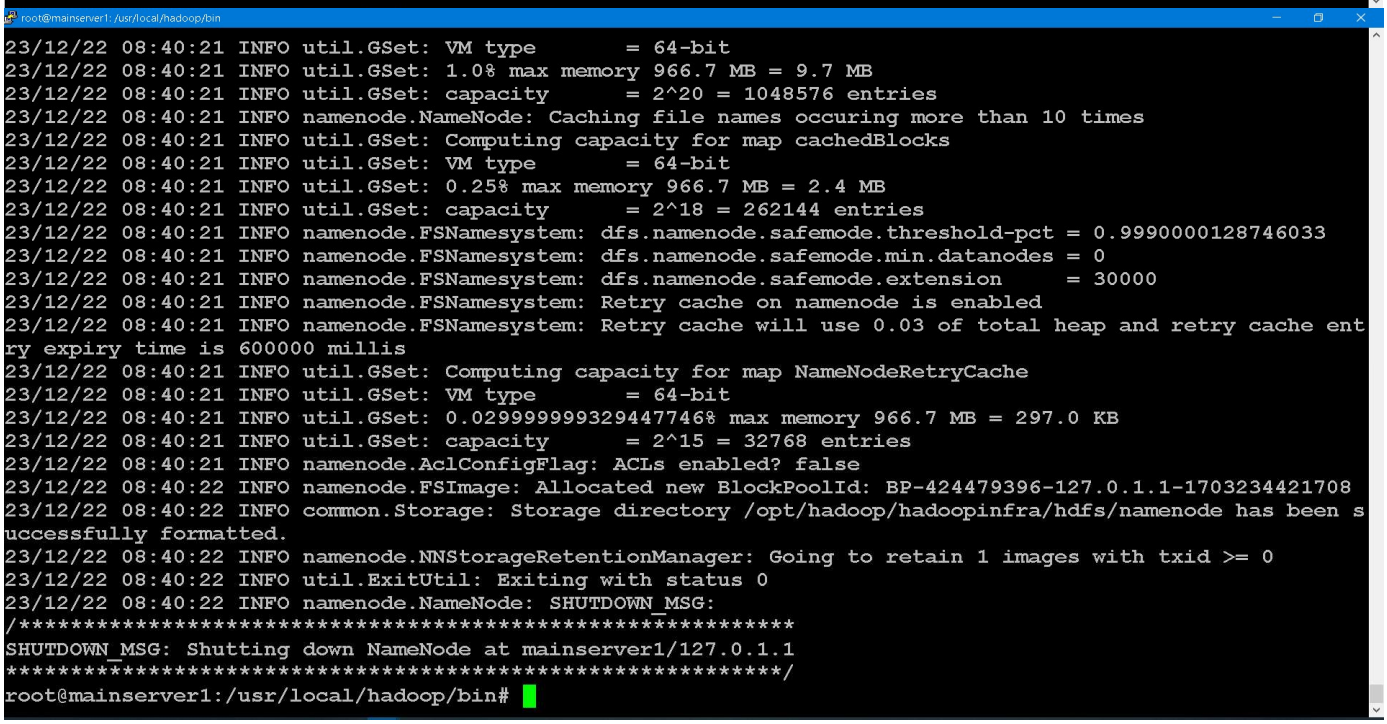
Edit hadoop-env.sh and add the following entry - on all the nodes

 # 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               mapred-site.xml
configuration.xsl           httpfs-env.sh               mapred-site.xml.template
container-executor.cfg      httpfs-log4j.properties     slaves
core-site.xml               httpfs-signature.secret     ssl-client.xml.example
hadoop-env.cmd              httpfs-site.xml             ssl-server.xml.example
hadoop-env.sh               log4j.properties            yarn-env.cmd
hadoop-metrics2.properties  mapred-env.cmd              yarn-env.sh
hadoop-metrics.properties   mapred-env.sh               yarn-site.xml
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#
```

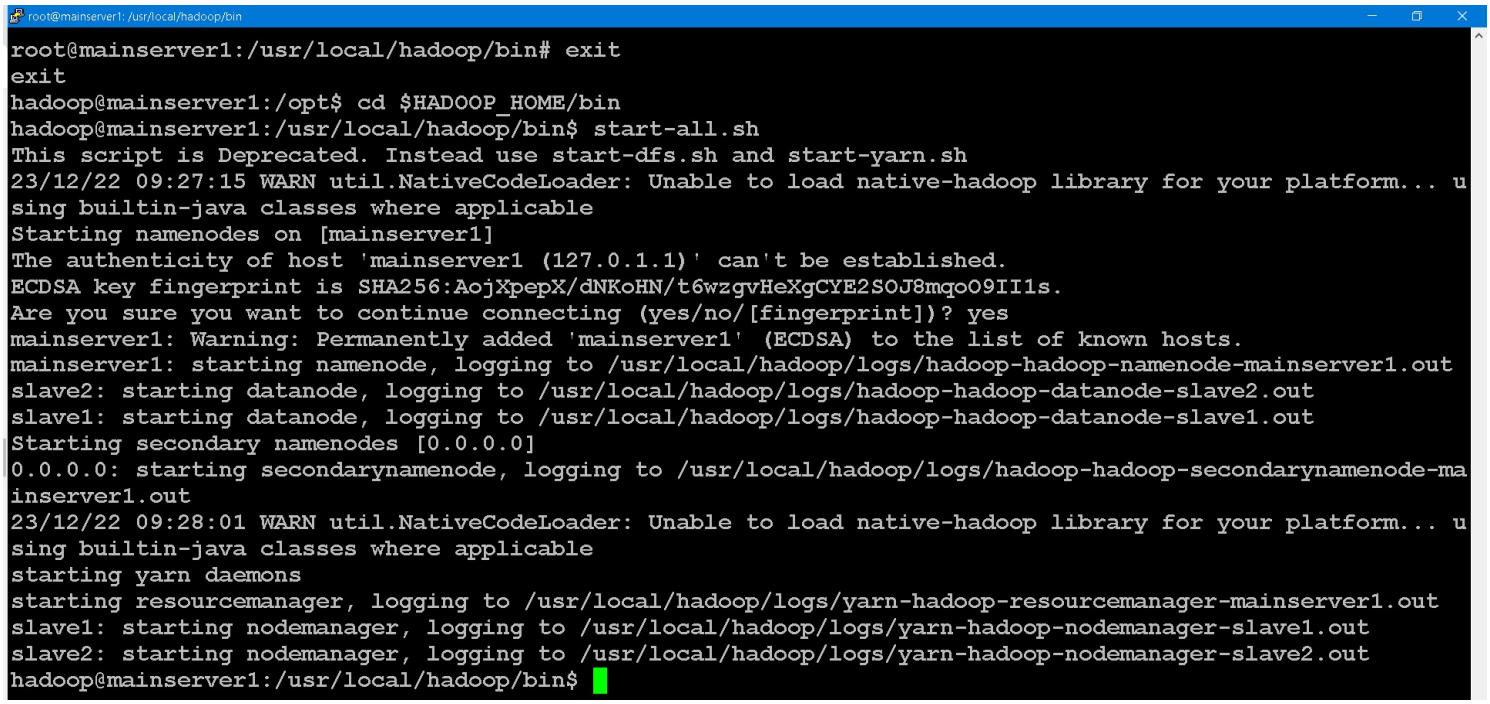
Execute the following command only on mainnode ie., master node only

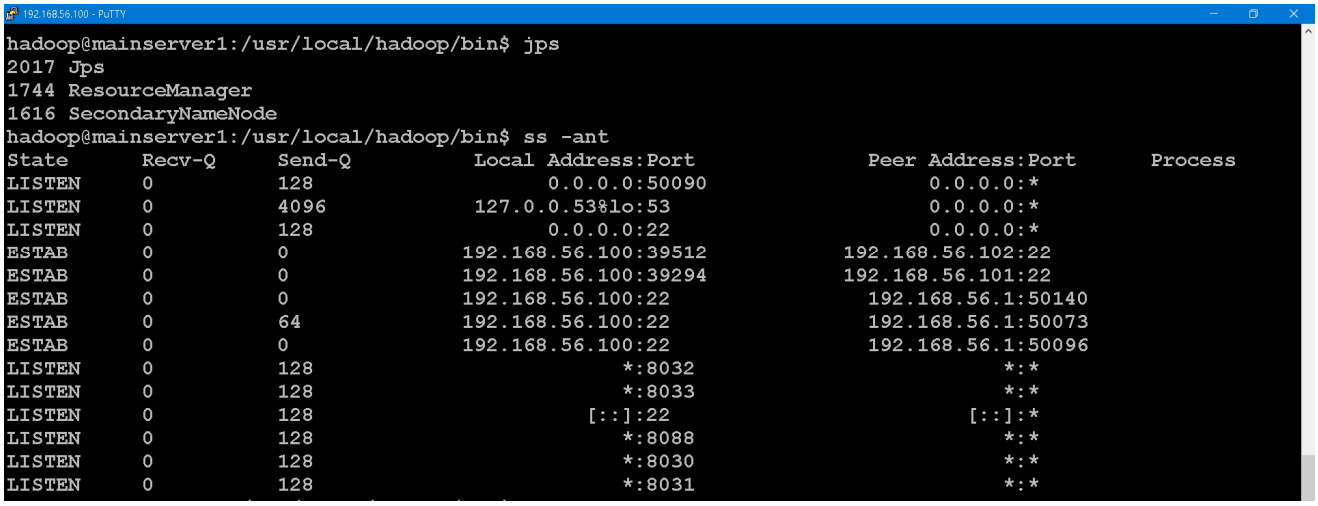


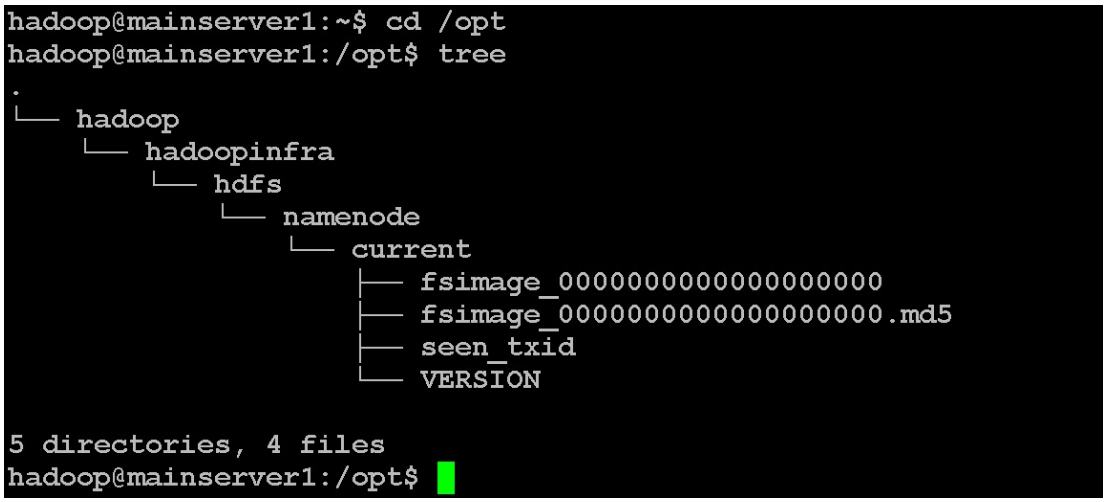


Note: We can also run the following command $ hdfs namenode -format

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


root@mainserver1: /usr/local/hadoop/bin
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/t6wzgvHeXgCYE2SOJ8mqo09II1s.
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_0000000000000000000
├── fsimage_0000000000000000000.md5
├── seen_txid
└── VERSION
5 directories, 4 files
hadoop@mainserver1:/opt$

```
hadoop@slave1: /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: /opt
hadoop@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#
```