

```
admin1@admin1-Lenovo-S510:~$ cd
Downloads
admin1@admin1-Lenovo-
S510:~/Downloads$ chmod 400
centos7.pem
admin1@admin1-Lenovo-
S510:~/Downloads$ ssh -i "centos7.pem"
centos@ec2-44-206-250-247.compute-
1.amazonaws.com
The authenticity of host 'ec2-44-206-250-
247.compute-1.amazonaws.com
(44.206.250.247)' can't be established.
ECDSA key fingerprint is
SHA256:yXHbAr+hApXJ49V7THmud+E
7hGUsPwLS5KCmJpS8IbL.
Are you sure you want to continue
connecting (yes/no)?
yes
```

```
Warning: Permanently added 'ec2-44-206-
250-247.compute-
1.amazonaws.com,44.206.250.247'
(ECDSA) to the list of known hosts.
```

## 1. Java Installation :

```
[centos@ip-172-31-86-192 ~]$sudo yum
install java-1.8.0-openjdk-devel
```

```
[centos@ip-172-31-86-192 ~]$java -version

openjdk version "1.8.0_312"

OpenJDK Runtime Environment (build
1.8.0_312-b07)

OpenJDK 64-Bit Server VM (build
25.312-b07, mixed mode)
```

```
[centos@ip-172-31-86-192 ~]$ readlink -f
$(which java)

/usr/lib/jvm/java-1.8.0-openjdk-
1.8.0.362.b08-1.el7_9.x86_64/jre/bin/java
```

```
[centos@ip-172-31-86-192 ~]$sudo vi
/etc/profile
```

(Add below 2 lines at end)

```
export
JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-
1.8.0.362.b08-1.el7_9.x86_64/jre
```

```
export
PATH=$JAVA_HOME/bin:$PATH
```

ctr+x -----> y----> enter (save and exit)

\$source /etc/profile (to execute the changes)

\$echo \$JAVA\_HOME

## 2. Passwordless communication

```
ssh-keygen
```

```
cd .ssh/
```

```
cat id_rsa.pub >>authorized_keys
```

```
ssh localhost
```

## 3. Download And Extract Hadoop

```
wget
https://archive.apache.org/dist/hadoop/core/
hadoop-1.2.1/hadoop-1.2.1.tar.gz
```

```
tar -xf hadoop-1.2.1.tar.gz
```

```
sudo mv hadoop-1.2.1 /usr/local/hadoop
```

## 4.Update bashrc file

```
[centos@ip-172-31-86-192 ~]$ sudo vi .bashrc
```

(add following at the end of .bashrc file)

```
export
HADOOP_PREFIX=/usr/local/hadoop/

export
PATH=$PATH:$HADOOP_PREFIX/bin

export JAVA_HOME=
/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.362.b08-
1.el7_9.x86_64/jre      export
PATH=$PATH:$JAVA_HOME/bin
```

ctr+x -----> y----> enter (save and exit)

exec bash or bash (to execute bashrc file)

-----  
-----

#### 5.Update hadoop-env.sh

```
[centos@ip-172-31-86-192 ~]$cd
/usr/local/hadoop/

[centos@ip-172-31-86-192 hadoop]$cd conf/

[centos@ip-172-31-86-192 conf]$sudo vi
hadoop-env.sh
```

(Add following 2 lines)

```
export JAVA_HOME=
/usr/lib/jvm/java-1.8.0-openjdk-
1.8.0.362.b08-1.el7_9.x86_64/jre
export HADOOP_OPTS=-
Djava.net.preferIPv4Stack=true
```

-----  
-----

#### 6.Namenode configure

```
#core-site.xml

[centos@ip-172-31-86-192 ~]$cd
/usr/local/hadoop/

[centos@ip-172-31-86-192 hadoop]$cd conf/

[centos@ip-172-31-86-192 conf]$sudo vi
core-site.xml
```

<configuration>

<property>

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

<value>hdfs://localhost:9000</value>

</property>

<property>

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

<value>/usr/local/hadoop/tmp</valu  
e>

</property>

</configuration>

#create "tmp" directory

```
[centos@ip-172-31-86-192 ~]$ mkdir -p
/usr/local/hadoop/tmp
```

-----  
-----

#### 7.Configure Datanode

#hdfs-site.xml

```
[centos@ip-172-31-86-192 ~]$cd
/usr/local/hadoop/

[centos@ip-172-31-86-192 hadoop]$cd conf/

[centos@ip-172-31-86-192 conf]$sudo vi
hdfs-site.xml
```

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.permissions</name>

```

<value>false</value>

</property>

</configuration>
-----
</property>
-----
</configuration>
-----
8.Configure Jobtracker/Map-Reduce
#mapred-site.xml

[centos@ip-172-31-86-192 ~]$cd
/usr/local/hadoop/

[centos@ip-172-31-86-192 hadoop]$cd conf/

[centos@ip-172-31-86-192 conf]$sudo vi
mapred-site.xml

<configuration>

    <property>

</property>

</configuration>
-----
9.exec bash

10.hadoop namenode -format

11.start-dfs.sh

12.start-mapred.sh

#### start-all.sh (alternate to step 12 & 13)

13.jps

```

## OUTPUT

```

Activities  Terminal  Wed 14:49  centos@ip-172-31-87-144:~
File Edit View Search Terminal Help
23/05/17 09:07:41 INFO util.GSet: Computing capacity for map BlocksMap
23/05/17 09:07:41 INFO util.GSet: VM type = 64-bit
23/05/17 09:07:41 INFO util.GSet: 2.0% max memory = 1013645312
23/05/17 09:07:41 INFO util.GSet: capacity = 2^21 = 2097152 entries
23/05/17 09:07:41 INFO util.GSet: recommended=2097152, actual=2097152
23/05/17 09:07:41 INFO namenode.FSNamesystem: fsOwner=centos
23/05/17 09:07:41 INFO namenode.FSNamesystem: supergroup=supergroup
23/05/17 09:07:41 INFO namenode.FSNamesystem: isPermissionEnabled=false
23/05/17 09:07:41 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
23/05/17 09:07:41 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessTokenLifetime=0 min(s), accessTokenLifetime=0 min(s)
23/05/17 09:07:41 INFO namenode.FSEditLog: dfs.namenode.edits.tolerant.length = 0
23/05/17 09:07:41 INFO namenode.NameNode: Caching file names occurring more than 10 times
23/05/17 09:07:42 INFO common.Storage: Image file /usr/local/hadoop/tmp/dfs/name/current/fsimage of size 112 bytes saved in 0 seconds.
23/05/17 09:07:42 INFO namenode.FSEditLog: closing edit log: position=4, editlog=/usr/local/hadoop/tmp/dfs/name/current/edits
23/05/17 09:07:42 INFO namenode.FSEditLog: close success: truncate to 4, editlog=/usr/local/hadoop/tmp/dfs/name/current/edits
23/05/17 09:07:42 INFO common.Storage: Storage directory /usr/local/hadoop/tmp/dfs/name has been successfully formatted.
23/05/17 09:07:42 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-87-144.ec2.internal/172.31.87.144
*****/
[centos@ip-172-31-87-144 conf]$ start-all.sh
starting namenode, logging to /usr/local/hadoop/libexec/../logs/hadoop-centos-namenode-ip-172-31-87-144.ec2.internal.out
localhost: starting datanode, logging to /usr/local/hadoop/libexec/../logs/hadoop-centos-datanode-ip-172-31-87-144.ec2.internal.out
localhost: starting secondarynamenode, logging to /usr/local/hadoop/libexec/../logs/hadoop-centos-secondarynamenode-ip-172-31-87-144.ec2.internal.out
starting jobtracker, logging to /usr/local/hadoop/libexec/../logs/hadoop-centos-jobtracker-ip-172-31-87-144.ec2.internal.out
localhost: starting tasktracker, logging to /usr/local/hadoop/libexec/../logs/hadoop-centos-tasktracker-ip-172-31-87-144.ec2.internal.out
[centos@ip-172-31-87-144 conf]$ jps
2241 NameNode
2498 SecondaryNameNode
2722 TaskTracker
2802 Jps
2361 DataNode
2590 JobTracker
[centos@ip-172-31-87-144 conf]$ ls
capacity-scheduler.xml  fair-scheduler.xml  hadoop-policy.xml  mapred-queue-acls.xml  slaves  taskcontroller.cfg
configuration.xml      hadoop-env.sh      hdfs-site.xml     mapred-site.xml       ssl-client.xml.example  task-log4j.properties
core-site.xml          hadoop-metrics2.properties  log4j.properties  masters               ssl-server.xml.example
[centos@ip-172-31-87-144 conf]$ cd
[centos@ip-172-31-87-144 ~]$ ls
hadoop-1.2.1.tar.gz
[centos@ip-172-31-87-144 ~]$ ls
hadoop
[centos@ip-172-31-87-144 ~]$ ls
chatbot.py  hadoop-1.2.1.tar.gz
[centos@ip-172-31-87-144 ~]$

```

Activities Firefox Web Browser Wed 14:38

Instance details | EC2 M... x Hadoop NameNode localho... x /usr/lib/jvm/jav/usr/lib/jv... x +

44.206.250.247:50070/dfshealth.jsp

## NameNode 'localhost:9000'

**Started:** Wed May 17 09:07:52 UTC 2023  
**Version:** 1.2.1, r1503152  
**Compiled:** Mon Jul 22 15:23:09 PDT 2013 by mattf  
**Upgrades:** There are no upgrades in progress.

[Browse the filesystem](#)  
[Namenode Logs](#)

### Cluster Summary

8 files and directories, 1 blocks = 9 total. Heap Size is 23.57 MB / 966.69 MB (2%)

Configured Capacity	: 9.99 GB
DFS Used	: 8.01 KB
Non DFS Used	: 2.01 GB
DFS Remaining	: 7.98 GB
DFS Used%	: 0 %
DFS Remaining%	: 79.86 %
<a href="#">Live Nodes</a>	: 1
<a href="#">Dead Nodes</a>	: 0
<a href="#">Decommissioning Nodes</a>	: 0
Number of Under-Replicated Blocks	: 0

### NameNode Storage:

Storage Directory	Type	State
/usr/local/hadoop/tmp/dfs/name	IMAGE_AND_EDITS	Active

44.206.250.247:50070/nn\_browsedfscontent.jsp

Activities Firefox Web Browser Wed 14:45

Instance details | EC2 M... x HDFS:/ x /usr/lib/jvm/jav/usr/lib/jv... x +

44.206.250.247:50075/browseDirectory.jsp?namenodeinfoPort=50070&dir=

## Contents of directory /

Goto : / go

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
<a href="#">usr</a>	dir				2023-05-17 09:08	rw-r-xr-x	centos	supergroup

[Go back to DFS home](#)

## Local logs

[Log](#) directory

This is [Apache Hadoop](#) release 1.2.1

Activities Firefox Web Browser Wed 14:45

Instance details | EC2 M... X HDFS:/usr/local/hadoop X /usr/lib/jvm/jav/usr/lib/jv X +

44.206.250.247:50075/browseDirectory.jsp?dir=%2Fusr%2Flocal%2Fhadoop&namenodeInfoPort=50070

### Contents of directory /usr/local/hadoop

Goto : /usr/local/hadoop go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
tmp	dir				2023-05-17 09:08	rw-r-xr-x	centos	supergroup

[Go back to DFS home](#)

### Local logs

[Log](#) directory

This is [Apache Hadoop](#) release 1.2.1

Activities Firefox Web Browser Wed 14:45

Instance details | EC2 M... X HDFS:/usr/local/hadoop/tmp X /usr/lib/jvm/jav/usr/lib/jv X +

44.206.250.247:50075/browseDirectory.jsp?dir=%2Fusr%2Flocal%2Fhadoop%2Ftmp%2Fmapred&namenodeInfoPort=50070

### Contents of directory /usr/local/hadoop/tmp/mapred

Goto : /usr/local/hadoop/tmp/mapred go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
system	dir				2023-05-17 09:08	rw-x-----	centos	supergroup

[Go back to DFS home](#)

### Local logs

[Log](#) directory

This is [Apache Hadoop](#) release 1.2.1



