

Installing HDFS on EC2:-

installing java

```
wget --no-check-certificate --no-cookies --header "Cookie:
oraclelicense=accept-securebackup-cookie" http://download.oracle.com/otn-
pub/java/jdk/8u141-b15/336fa29ff2bb4ef291e347e091f7f4a7/jdk-8u141-linux-
x64.rpm
```

```
sudo yum install -y jdk-8u141-linux-x64.rpm
```

Java -version

Installing Hadoop

```
wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/
hadoop-2.7.3.tar.gz
tar -xvf hadoop-2.7.3.tar.gz
```

```
## config bashrc
vi ~/.bashrc
```

Hadoop Properties

```
export HADOOP_HOME=$HOME/hadoop-2.7.3
export HADOOP_COMMON_HOME=$HOME/hadoop-2.7.3
export HADOOP_CONF_DIR=$HOME/hadoop-2.7.3/etc/hadoop
export HADOOP_HDFS_HOME=$HOME/hadoop-2.7.3
export HADOOP_MAPRED_HOME=$HOME/hadoop-2.7.3
export YARN_HOME=$HOME/hadoop-2.7.3
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
export PATH=$PATH:$HOME/hadoop-2.7.3/bin
export HADOOP_PID_DIR=$HOME/hadoop-2.7.3/hadoop2_data/hdfs/pid
```

```
export PATH=$PATH:$HOME/hadoop-2.7.3/bin
```

```
export JAVA_HOME=/usr/java/jdk1.8.0_141
export PATH=/usr/java/jdk1.8.0_141/bin:$PATH
```

Set HIVE_HOME

```
export HIVE_HOME=/root/apache-hive-2.1.0-bin
export PATH=$PATH:/root/apache-hive-2.1.0-bin/bin
```

```
source ~/.bashrc
```

Change conf/core-site.xml to

```
<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

Change conf/hdfs-site.xml to

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

Change conf/mapred-site.xml to

cp mapred-site.xml.template mapred-site.xml

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:9001</value>
</property>
</configuration>
```

Edit conf/hadoop-env.sh, look for JAVA_HOME, and set it up
export JAVA_HOME=/usr/java/jdk1.8.0_141

Setup passwordless ssh with the following steps:

```
ssh-keygen -t dsa -P '' -f ~/.ssh/id_dsa
cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
```

```
chmod -R 700 ~/.ssh
chmod 600 ~/.ssh/*
```

in file /etc/ssh/sshd_config, make sure that line containing
"PasswordAuthentication yes" is uncommented.
systemctl restart ssh

Ssh localhost

```
## then
cd hadoop-2.7.3
bin/hadoop namenode -format
cd hadoop-2.7.3/sbin
./start-all.sh
```

```
jps
```

=====line =====

Installing Hive:

```
## Download Hive tar.
```

```
wget http://archive.apache.org/dist/hive/hive-2.1.0/apache-hive-2.1.0-bin.tar.gz
tar -xzf apache-hive-2.1.0-bin.tar.gz
ls
```

```
## check version
Hive --version
```

Following by,

<https://www.edureka.co/blog/apache-hive-installation-on-ubuntu>

References:

<https://blog.knoldus.com/installing-latest-oracle-jdk-on-linux-ec2-instance-centos/>

<https://www.codementor.io/@lakshaynagpal/how-to-setup-hadoop-pseudo-distributed-mode-single-cluster-du10831te>

<https://www.edureka.co/blog/install-hadoop-single-node-hadoop-cluster>

<https://gist.github.com/samarthbhargav/b522819b92a66a0c3afb>

<https://stackoverflow.com/questions/4594048/how-do-i-setup-passwordless-ssh-on-aws>

<https://www.edureka.co/blog/apache-hive-installation-on-ubuntu>