

- a. Java is installed with the command `$ sudo apt-get install openjdk-7-jdk`, we need to update JAVA\_HOME in `.bashrc`

```
# Set JAVA_HOME
export JAVA_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64
export PATH=$PATH:$JAVA_HOME/bin
```

- b. Openssh is already installed, run the commands to make it passphrase less

```
$ssh-keygen -t rsa -P ""
$cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
$ssh-add
```

## 1. Download and Unpack Hadoop

- a. Navigate to Downloads and extract **hadoop-2.7.1.tar.gz**
- b. Move the **hadoop-2.7.1** directory to `/home/user1/HadoopInstallations`
- c. Update Hadoop Environment variables in `.bashrc`

```
# Set HADOOP_HOME
export HADOOP_HOME=/home/user1/HadoopInstallations/hadoop-2.7.1
export HADOOP_HDFS_HOME=/home/user1/HadoopInstallations/hadoop-2.7.1
export HADOOP_COMMON_HOME=/home/user1/HadoopInstallations/hadoop-2.7.1
export HADOOP_MAPRED_HOME=/home/user1/HadoopInstallations/hadoop-2.7.1
export YARN_HOME=/home/user1/HadoopInstallations/hadoop-2.7.1
export HADOOP_CONF_DIR=/home/user1/HadoopInstallations/hadoop-2.7.1/etc/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
```

## 2. Customize Hadoop (Files in `$HADOOP_HOME/etc/hadoop` directory)

- a. Update the properties in `core-site.xml`

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

  <property>
    <name>hadoop.tmp.dir</name>
    <value>/home/user1/tmp</value>
  </property>
</configuration>
```

- b. Update the properties in `hdfs-site.xml`

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

  <property>
    <name>dfs.permissions</name>
    <value>>false</value>
  </property>
</configuration>
```

- c. Update the properties in mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

- d. Update the properties in yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>

  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

- e. Update the JAVA\_HOME in hadoop-env.sh

```
export JAVA_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64
```

3. Format the NameNode

```
$ hdfs namenode -format
$ hdfs datanode -format (Optional for our installation)
```

4. Start Hadoop Services

```
$ cd $HADOOP_HOME/sbin
$ ./start-all.sh
```