1. Install Java 8 with following commands:

```
sudo add-apt-repository ppa:webupd8team/java
sudo apt-get update
sudo apt-get install oracle-java8-installer
```

2. After installation, make a quick check whether Sun's JDK is correctly set up:

```
java -version
```

3. Configure java environment

```
sudo apt-get install oracle-java8-set-default
```

4. Install ssh

```
sudo apt-get install openssh-server
```

5. Disable ipv6 by adding following to /etc/sysctl.conf

```
net.ipv6.conf.all.disable ipv6 = 1
net.ipv6.conf.default.disable ipv6 = 1
net.ipv6.conf.lo.disable ipv6 = 1
```

6. Creating a hduser user for accessing HDFS and MapReduce

```
sudo adduser hduser
```

7. Download hadoop-2.9.1.tar.gz source from Apache mirrors (http://www-eu.apache.org/dist/hadoop/common/stable/)

```
## Locate to hadoop installation parent dir
     cd /usr/local/

## Extract Hadoop source
     sudo tar -xzvf hadoop-2.9.1.tar.gz

## Move hadoop-2.9.1 to hadoop folder
     sudo mv hadoop-2.9.1 /usr/local/hadoop

## Assign ownership of this folder to Hadoop user
     sudo chown hduser -R /usr/local/hadoop
```

8. Configuring SSH . First login with hduser (**and from now use only hduser account for further steps)**

```
     sudo su hduser

## Generate ssh key for hduser account
     ssh-keygen -t rsa -P ""


## Copy id rsa.pub to authorized keys from hduser
     cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

9. Check if ssh working by following command:

```
ssh localhost
exit
```

10. Add following to ~/.bashrc

```
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$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 PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

11. Nexxt run bashrc

source ~/.bashrc

12. Now edit $HADOOP HOME/etc/hadoop/hadoop-env.sh file and set JAVA_HOME environment variable. Change the JAVA path as per install on your system. This path may vary as per your operating system version and installation source. So make sure you are using correct path.

```
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
```

13. Setup Hadoop Configuration Files(the files are in /usr/local/hadoop/etc/hadoop directory)

(a) Create namenode and datanode directory with following command:

```
mkdir -p /usr/local/hadoop/hadoopdata/hdfs/namenode
mkdir -p /usr/local/hadoop/hadoopdata/hdfs/datanode
```

(b) Edit /usr/local/hadoop/etc/hadoop/core-site.xml. Add following inside the configuration tag

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

(c) Edit /usr/local/hadoop/etc/hadoop/hdfs-site.xml . Add following inside the configuration tag

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


<property>
     <name>dfs.name.dir</name>
     <value>file:///usr/local/hadoop/hadoopdata/
     hdfs/namenode</value>
```

```
</property>

<property>
     <name>dfs.data.dir</name>
     <value>file:///usr/local/hadoop/
     hadoopdata/hdfs/datanode</value>
</property>
```

(d) Edit /usr/local/hadoop/etc/hadoop/mapred-site.xml. If mapred.site.xml is not there then create using following command:

```
cp  /usr/local/hadoop/etc/hadoop/mapred.site.xml.template
     /usr/local/hadoop/etc/hadoop/mapred.site.xml
```

Now add following inside the configuration tag

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


<property>
     <name>yarn.app.mapreduce.am.env</name>
     <value>HADOOP_MAPRED_HOME=/usr/local/hadoop/</value>
</property>

<property>
     <name>mapreduce.map.env</name>
     <value>HADOOP_MAPRED_HOME=/usr/local/hadoop/</value>
</property>

<property>
     <name>mapreduce.reduce.env</name>
     <value>HADOOP_MAPRED_HOME=/usr/local/hadoop/</value>
</property>
```

(e) Edit yarn-site.xml by adding following inside the configuration tag

```
<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>
```

14. Format namenode

```
hdfs namenode -format
```

15. Ensure that clusterId in /usr/local/hadoop/hadoopdata/hdfs/datanode/current/VERSION file a and clusterID in /usr/local/hadoop/hadoopdata/hdfs/namenode/current/VERSION are same. If they are not same then copy the clusterId from /usr/local/hadoop/hadoopdata/ hdfs/datanode/current/V ERSION and reset it in /usr/local/hadoop/hadoopdata/ hdfs/namenode/current/VERSION

16. Start all the deamons usinf following command

```
start-all.sh
```

17. Check if all the deamons are running with following command

```
jps
```

16. Run Wordcount example following commands:

```
hadoop fs -mkdir -p /user/hduser
hadoop fs -put file.txt file
Compile the .java files
javac -cp 'hadoop classpath' *.java
Create the jar
jar cvf test.jar *.class
```

17. Running the program

```
hadoop jar test.jar WordCount file testoutput
```