# Setting up a Hadoop cluster

## In all the nodes do the following steps;

- vim /etc/hosts ---> give the ip's and namenodes in this for all the slaves, master and the clients.
- rpm –i –v –h jdk-8u171-linux-X64.rpm
- cd /usr/java/ jdk1.8.0-171-amd64/
- JAVA_HOME= /usr/java/ jdk1.8.0_171-amd64/
- echo $JAVA_HOME
- echo $PATH
- vi /root/.bashrc

```
export JAVA_HOME= /usr/java/ jdk1.8.0_171-amd64
export PATH= /usr/java/ jdk1.8.0_171-amd64/bin:$PATH
```

- rpm –ivh hadoop-1.2.1-1.x86-64.rpm –force
- hadoop version

## Master Setup:

let the ip is 192.168.43.200

- mkdir /master
- vim hdfs-site.xml

```
<configuration>
<property>
<name>dfs.name.dir</name>
<value>/master</value>
</property>
</configuration>
:wq
```

- cd /etc/hadoop
- vi core-site.xml

```
<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://192.168.43.200:9001</value>
</property>
</configuration>
```

```
:wq
```

- `hadoop namenode –format`
  ```
  -Y
  ```
- `hadoop-daemon.sh start namenode`
- `jps`
- `hadoop dfsadmin –report`
- `systemctl stop firewalld`
- `systemctl disable firewalld`

Go to the web browser and open the link http://192.168.43.200:50070

## Slave Setup:

- `mkdir /data`
- `cd /etc/hadoop`
- `vi hdfs-site.xml`
  ```
  <configuration>
  <property>
  <name>dfs.data.dir</name>
  <value>/data</value>
  </property>
  </configuration>
  :wq
  ```
- `cd /etc/hadoop/`
- `vi core-site.xml`
  ```
  <configuration>
  <property>
  <name>fs.default.name</name>
  <value>hdfs://192.168.43.200:9001</value>
  </property>
  </configuration>
  :wq
  ```
- `hadoop-daemon.sh  start datanode`
- `jps`
- `systemctl stop firewalld`
- `systemctl disable firewalld`

In case of any error just remove the file /data and make it again and then start the service.

## Client Setup:

- cd /etc/hadoop
- vi core-site.xml

```
<configuration>
 <property>
 <name>fs.default.name</name>
 <value>hdfs://192.168.43.200:9001</value>
 </property>
 </configuration>
 :wq
```

- hadoop fs –ls /
- hadoop fs –mkdir /dir
- hadoop fs –touchz /my.txt
- vim filename.txt

```
 adding data to this file to be uploaded on the cluster
```

- hadoop fs –put filename.txt /
- hadoop fs –cat /filename.txt