

## Hadoop setup for windows 10 word count

Hadoop Version used 3.3.1 with available bin of 3.3.0

To setup Hadoop in windows follow link in next cell

<https://youtu.be/GfixwKmS8Rg>

After setting up hadoop in windows 10 use start-all.cmd to start all deamons

Start your cmd as admin to avoid daemon shutdown due to permissions

1. First make a Folder Wordcount -> Create file [mapper.py](#) and [reducer.py](#) with required python code.
2. Then put your text file in hdfs using -put command.
3. Then we have to change mapred-site.xml in etc/hadoop. It is necessary to use mapreduce. place below configuration inside mapred-site.xml

4. Configuration for mapred-site.xml properties

```
<configuration>
<property>
<name>yarn.app.mapreduce.am.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>mapreduce.map.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>mapreduce.reduce.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
</configuration>
```

5. Commands to run word-count. First,cd to folder containing mapper.py and reducer.py else you should pass full path to mapper.py and reducer.py and also for hadoop-streaming-3.3.1.jar

6. Command used in my cmd is mentioned below.

D:\MapReduce>hadoop jar hadoop-streaming-3.3.1.jar -mapper "python mapper1.py" -reducer "python reducer.py" -input /sample/story.txt -output /sample/result1

**My Directory was MapReduce inside MapReduce I have [mapper1.py](#) and [reducer.py](#) and hadoop-streaming-3.3.1.jar files. So, I am not giving full path to them in command.**

Where to find hadoop-streaming-3.3.1.jar. It is inside your C:\hadoop-3.3.1.tar\hadoop-3.3.1\share\hadoop\tools\lib

With all these settings and using above mentioned command. Hadoop streaming and map reduce will be succesfull. Most important is replacing configuration inside mapred-site.xml