

Running WordCount on Cloudera

1. Create a folder HADOOP/WC and HADOOP/files on the desktop using following command :

```
mkdir -p ~/Desktop/HADOOP/WC
mkdir -p ~/Desktop/HADOOP/files
```

2. Download Samplefile.txt and save it at ~/Desktop/HADOOP/files.

2. Next copy this file to the HDFS using following command :

```
hadoop fs -put ~/Desktop/HADOOP/files/Samplefile.txt file
```

3. Download WordCount.java file and save it at ~/Desktop/HADOOP/WC

Next compile it using following command:

```
cd ~/Desktop/HADOOP/WC/
javac -cp /usr/lib/hadoop/*:/usr/lib/hadoop-mapreduce/*
WordCount.java
```

4. Create a JAR file for the WordCount application.

```
jar cvf wordcount.jar *.class
```

5. Run the WordCount application from the JAR file, passing the paths to the input and output directories in HDFS.

```
hadoop jar wordcount.jar WordCount file WOutput
```

6. To look at the output use the following command:

```
hadoop fs -cat WOutput/part*
```