

```
#word count using spark
```

```
1)make a folder DSBDA and enter the spark file in it
```

```
2)echo $SHELL
```

```
3)nano ~/.bashrc
```

```
4)enter the path of file in the export
```

```
PATH="$HOME/DSBDAL/sparkfilename/bin:$PATH"
```

```
5) source ~/.bashrc confirm the path
```

```
6)make two file one is save the input countword.txt and second  
wordcount_program.scala
```

```
val inputFile = sc.textFile("countword_input.txt")
```

```
val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,  
1)).reduceByKey(_ + _)
```

```
counts.toDebugString
```

```
counts.cache()
```

```
counts.saveAsTextFile("output")
```

```
this program type in program file
```

```
7)then type spark-shell < wordcount.scala
```

```
8)check the output cat output/part-*
```

```
9)if change in program then delete the output first then run the program  
rm -r output
```