

Week 3 commands

```
val datafile = sc.textFile("hdfs:///user/user1/textData/textData.txt")
```

```
datafile.take(20).foreach(println)
```

```
val words = datafile.flatMap(_.split(" "))
```

```
val lowerWords = words.map(_.toLowerCase())
```

```
lowerWords.toString
```

```
val romeo = lowerWords.filter(_=="romeo")
```

```
romeo.toString
```

```
romeo.count
```

```
val allwords = lowerWords.map((_,1))
```

```
val wordCounts = allwords.reduceByKey(_+_)
```

```
wordCounts.toString
```

```
wordCounts.take(10).foreach(println)
```

```
wordCounts.sortBy(_._1).take(10).foreach(println)
```

```
wordCounts.sortBy(_._1,false).take(10).foreach(println)
```

```
wordCounts.cache
```

```
wordCounts.toDebugString
```

```
val popularWords = wordCounts.filter(_._2>2500)
```

```
popularWords.count
```

```
popularWords.sortBy(_._2).collect.foreach(println)
```

```
val charCounts =  
popularWords.flatMap(_._1.toCharArray).map((_,1)).reduceByKey(_  
+_)
```

```
charCounts.count
```

```
charCounts.toDebugString
```

```
charCounts.collect.foreach(println)
```