

Assignment 13

Installing Java

```
sudo apt-get install openjdk-7-jdk
```

Installing Scala

```
sudo apt-get install scala
```

Installing Spark

```
tar -xvf spark-2.4.1-bin-hadoop2.7.tgz
```

Installing OpenSSH

```
sudo apt-get install openssh-server
```

Editing the bashrc file for spark shell

```
sudo nano ~/.bashrc
```

```
export PATH = $PATH: /usr/local/spark/bin
```

Running the spark shell

```
spark-shell
```

Understanding Word Count Example in Scala

```
package com.oreilly.learningsparkexamples.mini.scala
import org.apache.spark._
import org.apache.spark.SparkContext._
object WordCountObject {
  def main(args: Array[String]) {
    val inputData = args(0)
    val outputData = args(1)
    val conf = new SparkConf().setAppName("wordCountExample")
    // Creating a sparkContext for app entry
    val sc = new SparkContext(conf)
    // Loading the input data
    val input = sc.textFile(inputData)
    // Splitting sentences into words based on empty spaces
    val all_words = input.flatMap(line => line.split(" "))
```

```
    // Transforming words word count key value pairs.  
    val wordCounts = all_words.map(word => (word,  
1)).reduceByKey{case (x, y) => x + y}  
    // Logging the results to the output file  
    counts.saveAsTextFile(outputData)  
  }  
}
```