

# CS5542

## Big Data Apps and Analytics

### LAB ASSIGNMENT #1

---

Student id: 19  
Name: harshil patel

Partner id: 5  
Name: santhosh

## Code:

```
17
18 val input=sc.textFile("u.data")
19
20 //val map = Map(keyValuePairs : *)
21 val i1=input.map(fields => fields.split("\t")(0) -> 1).cache() //this will create a key value pair consisting of userid and 1 (userid,1)
22
23 val ou2=i1.reduceByKey(_+_ ) //it will reduce the key value pair by key so getting as the count of the user as review a item.
24
25 val ou3=ou2.filter(x => {x._2 > 25}) //it will provide a filter to get only user that have reviewed more than 25 items.
26 ou3.saveAsTextFile("output3")
27 // val i=input.map(word=>(word.split(" ") (0),1)).cache()
28
```

After loading data into val. Input we will create key value pair of user id and 1 like (123,1) (234,1) and so on.

After this we will reduce key value pair so getting as the count of the user as review a item.

Than we will use filter to get only user that have reviewed more than 25 items.

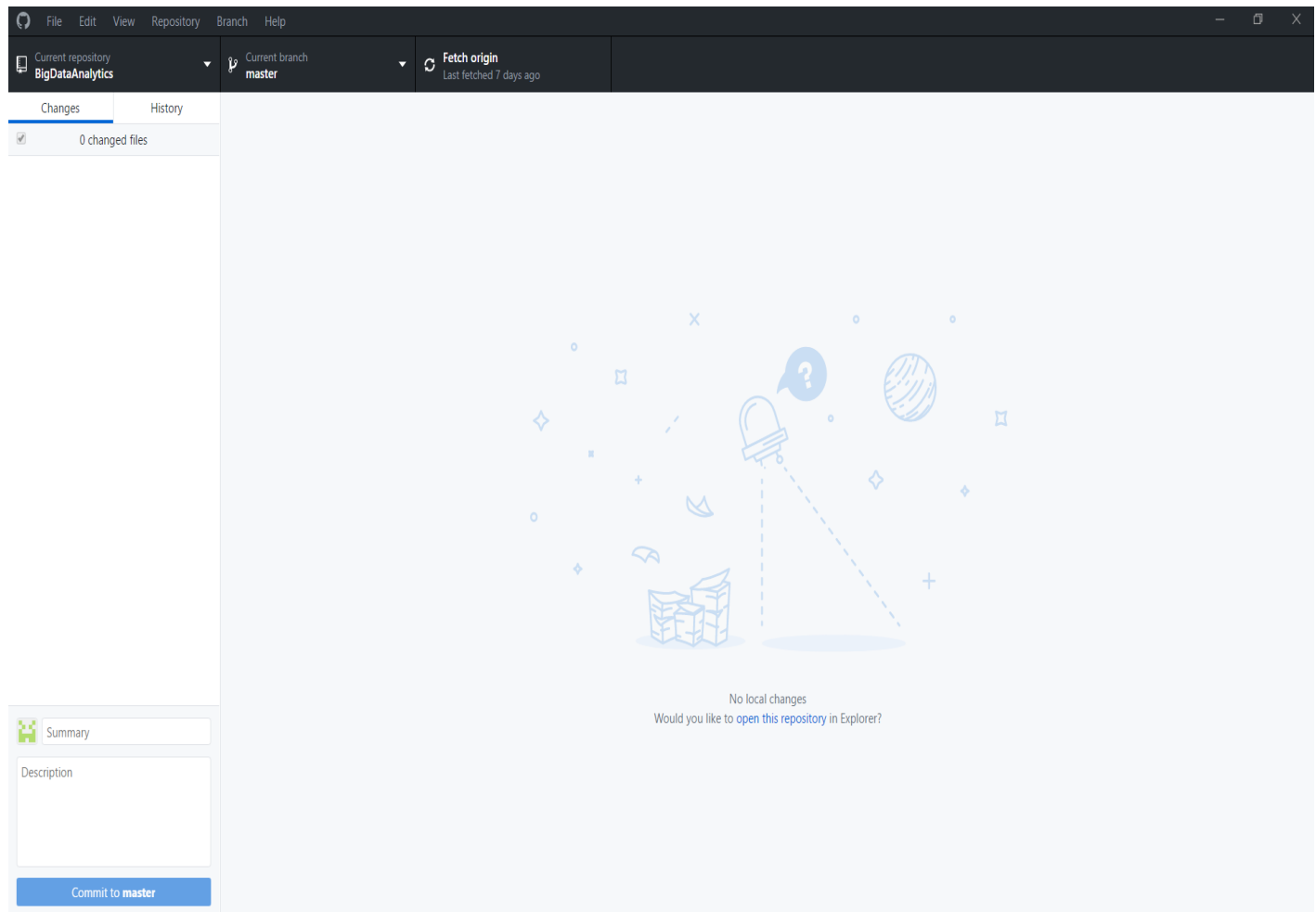
# Output:

The screenshot shows the IntelliJ IDEA IDE interface. The title bar indicates the project is 'CSS542-Tutorial18-SparkSourceCode' and the file being edited is 'SparkWordCount.scala'. The menu bar includes File, Edit, View, Navigate, Code, Analyze, Refactor, Build, Run, Tools, VCS, Window, and Help. The toolbar contains icons for running, debugging, and other IDE functions. The project explorer on the left shows the project structure, including folders for 'idea', 'output', 'output1', 'output2', 'output3', 'Spark WordCount [root]', 'SparkTransformationActions', 'input', 'u.data', 'u1', and 'External Libraries'. The main editor area displays the output of the Spark WordCount application, showing a list of word counts. The output is displayed in a table with two columns: a line number and a word count. The word counts are listed in parentheses, representing (word, count).

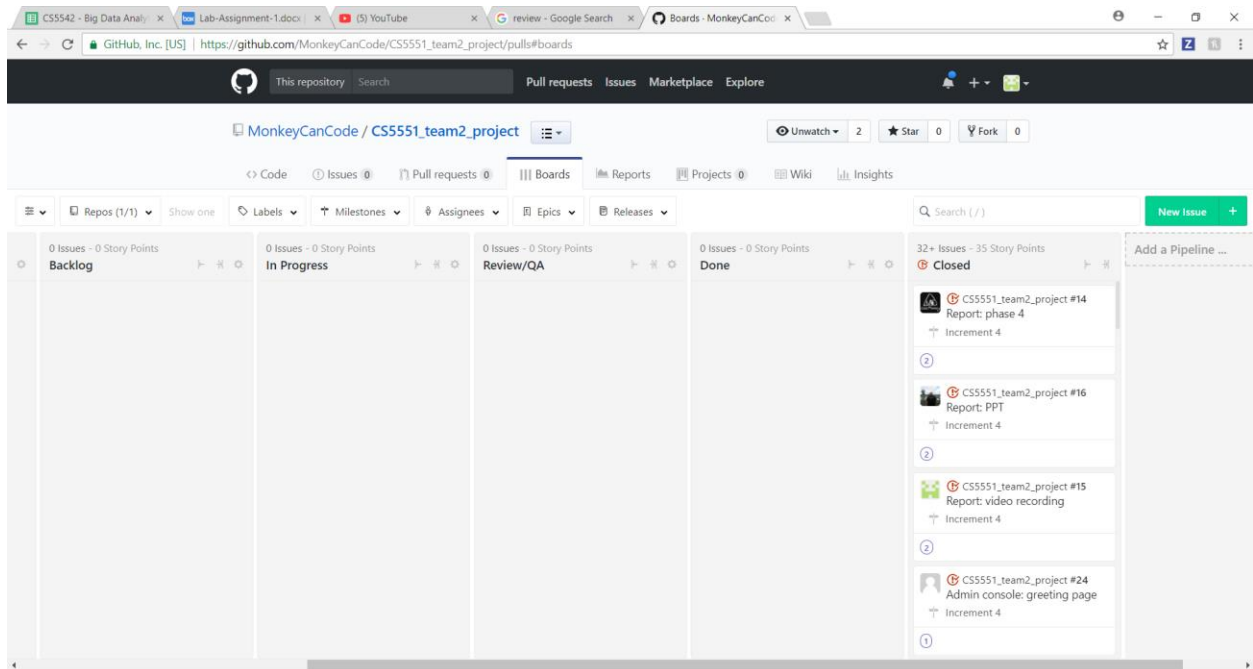
1	(253, 57)
2	(82, 168)
3	(426, 101)
4	(332, 183)
5	(927, 120)
6	(190, 55)
7	(312, 223)
8	(62, 232)
9	(664, 166)
10	(301, 275)
11	(734, 64)
12	(523, 100)
13	(275, 434)
14	(475, 202)
15	(875, 89)
16	(503, 160)
17	(846, 405)
18	(13, 636)
19	(642, 318)
20	(617, 108)
21	(774, 224)
22	(792, 47)
23	(853, 41)
24	(831, 73)
25	(466, 101)
26	(691, 32)
27	(815, 183)
28	(237, 48)
29	(101, 67)
30	(428, 55)
31	(112, 46)
32	(860, 71)
33	(550, 40)
34	(178, 273)
35	(745, 305)
36	(31, 36)
37	(71, 38)
38	(448, 38)
39	(462, 33)
40	(299, 280)
41	(556, 44)
42	(167, 65)

## 2. Github Desktop

---



# Github issues



This page consist of issue that I have created for my ASC projects.

This page show burndown chart.

