**CS5542 Big Data Apps and Analytics**

**LAB ASSIGNMENT #2**

**Name: Raj Anantharaman (#3)**

**Email:** [**rah59@mail.umkc.edu**](mailto:rah59@mail.umkc.edu)

**The spark program that I tried out for this assignment was to combine two inputs, the first is an array of customers and the second is an array of purchases that the customer has made before. The goal of the program is to combine these two inputs into a single RDD and then apply a transformation to add the purchase values. The RDD is finally saved to disk.**

**The following have been used in this program.**

Transformations: join, reduceByKey, sortByKey

Actions: count, cache, saveasTextFile

(1,10.25), (2,5.15),(3,2.35), (1,3.55), (2,1.25), (4,3.85), (4,2.50), (5,2.55), (5,1.25)

(1,(”raj”,13.80)), (2, (“arun”,5.15)), (3,(“George”,2.35)),…….

(1,(”raj”,13.80)), (2, (“arun”,5.15)), (3,(“George”,2.35)),…….

(1,(”raj”,10.25)), (1,(“raj”,3.55)), (2, (“arun”,5.15)), (3,(“George”,2.35)),…….

(1,(”raj”,10.25)), (3,(“George”,2.35)), (2,(“arun”,5.15)), (1,(“raj”,3.55)),…

Purchases

Customers

(1,”raj”), (2, “arun”), (3, “George”), (4,”Daniel”), (5,”vidya”)

Final Result

Reducing

Shuffling

Input

Combining

Code is available at <http://github.com/rah59/BigDataSpring2017> under Lab Assignment #2