***Double Secret Probation Assignment***

**Group Assignment, Assignment 4, Big Data, Spring 2019**

**Due: February 27, 2019 by end of class (or soon thereafter)**

Group Category B (Your group may not receive help from anyone outside your group.)

**Grade:** Will count as 5% of your overall course grade.

***Hint: This is very similar (but not exactly) the same problem as Parts 1 and 2 of your first individual assignment. The Python MapReduce solution and output is posted to the Assignment 1 folder.***

Part 1 and Part 2: Use DataA1.csv as the input. Data overview: The data is in a CSV file. The rows are Orders. The first column is Customer ID Number, the second column is Item ID Number, and the third column is Amount Spent on the Order. Note, that Customer ID Number and Item ID Number repeat (i.e., a customer will have multiple orders, and an item will be ordered more than once).

**Part 1:**

Create a Spark/Scala program to determine the Total Amount spent by Customer; ensure your output is in numerical order by customer ID (0, 1, 2, …). Execute. Submit your code (.scala file) and output (.txt file). Please see the Assignment 1 solution for output format.

**Part 2:**

Create a Spark/Scala program to sort the Total Amount spent by Customer (from the Customer who spent the most to the Customer who spent the least). **Only output the top 5 (i.e., the highest five, in order from highest to second highest, etc.).** Execute. Submit your code (.scala file) and output (.txt file). Please see the Assignment 1 solution for output format (other than what was mentioned above).

***Include the names of present team-members either at the top of the code files (commented out), or in a text file that you submit.***

**Submission:**

Please put all of your files into a single folder, ZIP it, and then submit via Blackboard. **Save your ZIP file as SectionNumber-TeamNumber.ZIP.** (If you have issues uploading to Blackboard, then email it to me directly. Email **both** my Mason and GMAIL.)

**Emails:** Joe.Wilck@mason.wm.edu Joe.Wilck@gmail.com

**Note:**

Students who do not attend will have an opportunity to “make-up” the assignment, but it will be an individual assignment (Category A) or a team made-up of people who missed the class (email me). Please do not include absent team-members on any communication.