**CS5542 Big Data Apps and Analytics**

**In Class Programming – 2**

**3rd September 2021 (11:59 pm CST)**

**Spark Programming:**

**Write a spark program to group the words in a given text file based on their starting letters.**

**Use the Text File provided with ICP (icp2.txt)**

Example Input Example Output

U, United, Unified

S, States, Separated

I, Incident, Investments

B, Board

United States Incident

Separated Unified

Investments Board

ICP Requirements:

1. Spark Integration with Colab (or IDE that you are using)
2. Creating a well commented Spark program and outputting the correct results and writing it to output file.
3. Code quality, wiki Report quality, video explanation

Submission Guidelines:

1. Sign in into your github account
2. Click on this link :<https://classroom.github.com/a/PWO-jM8z>
3. Accept the ICP-2 (Assignment 2)
4. Complete your ICP and create your wiki report (Pdf or Word doc).
5. Folders for ICP:
   1. Create two folders for source code and documentation
      1. Source code folder contains only the code and output file (if applicable)
      2. Documentation folder contains wiki report and the images of your results
6. The wiki report should have at least the following:
   1. What you learned in the ICP
   2. ICP description what was the task you were performing
   3. Challenges that you faced
   4. Screen shots that shows the successful execution of each required step of your code
   5. Out put file link if applicable
   6. Video link (YouTube or any other publicly available video platform)
   7. Any inside about the data or the ICP in general
7. Upload your ICP folders to your assignment GitHub repository
8. Click on the add a readme file on the next page and write done your name and email

You are all Done!!!!!!!