1. What you learned in the ICP

In this ICP I learned how to do Spark integration with Google Colab and read the data file using the Spark Context. In addition, I learned how to use split the data into words and remove all punctuations and stop-words.

1. ICP description what was the task you were performing

This task was about performing how to read your text file using spark as well as how to split the data and cluster each word that starts with the same alphabet together.

1. Challenges that you faced

The only challenge was how to divide the words based on the alphabet.

1. Screen shots that shows the successful execution of each required step of your code

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Video link:

<https://drive.google.com/file/d/15iKdJUgfFvcFV8cGS0biOjcEQGAGlCsY/view?usp=sharing>

1. Output file link if applicable

The output file has been uploaded to Github (under source code document)