**Big Data Analytics and Applications**

**In Class Programming 2**

1. What you learned in the ICP.

In this ICP I have learned to setup spark in Google Colab by importing and downloading all the required libraries and packages. And I have also learned how to print the data which is in a RDD Format and to group the words in a dataset based on their first character.

1. ICP description what was the task you were performing

In this ICP we are asked to setup spark in Google Colab or Pycharm and to write a spark program in order to group the words based on their first character.

1. Challenges that you faced

I have faced issues while I am setting up spark in Google Colab but I was able to resolve it.

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

These are the screenshots of the code which are used to import and download required packages and libraries in order to setup spark.

A screenshot of a social media post

Description automatically generated

These are the lines of code where I have imported Spark session and Spark Context and also to import the input file and display the data in the input file.

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

These are the lines of code to group the words in the input file based on their first character.

A close up of a logo

Description automatically generated

This is the screenshot of the output where the words with same character are grouped.

A screenshot of a cell phone

Description automatically generated

A close up of a logo

Description automatically generated

**YouTube Video Link:** <https://youtu.be/xdhc_EZI2P0>