

Lab 3

[Click here to Register Attendance](https://forms.gle/pesShFDe1ibVfS3L7)

|  |  |
| --- | --- |
| Name |  |
| Date |  |
| Student No |  |
| Student Email |  |

### **Running a MapReduce Job on your local machine**

1. Download the python file [here](https://github.com/marloftitsligo/ProgrammingForBigData) called MapReduceIris.py and the dataset file: Iris.data
2. Open a Command Line at that folder location (assuming you have Python installed)
3. Run the following command (replace iris.data with your data filename):  
   python ratingsbreakdown00.py iris.data
4. You will likely get an error message as this python script uses a library called mrjob.   
   To install the library type:  
   pip install mrjob  
   You should get a confirmation message: Successfully installed mrjob-0.7.4
5. Try running the original command again (replace iris.data with your data filename):  
   python ratingsbreakdown00.py iris.data
6. You should get the following in the output:   
   "setosa sepal width avg" 3.418
7. Post a screenshot of the output here:

|  |
| --- |
|  |

1. Try to adjust the Reducer calculation to find the average of other Species  
   Describe the rationales for your changes, your code and screenshot your output here:

|  |
| --- |
|  |

1. Go to the UCI Machine Learning Data Repository and find another dataset to use with this MapReduce code and make necessary changes to the code to perform an analysis of your choice.   
   Insert the name and a link to the dataset here and a short description of the analyses you performed:

|  |
| --- |
|  |

1. Push the code and screenshots from step 11 above to Github and post the link here:

|  |
| --- |
|  |