BerkeleyX: CS190.1x Scalable Machine Learning

LAB2: WORD COUNT (100/100 points)

Once you have successfully run the test notebook, you can submit to the course autograder by first **exporting lab2 as a Python (.py) file**, and then using the file chooser to select your file and clicking *"Check"* to submit your code.

Note that if you took CS100.1x, lab 1 (lab1_word_count_student.ipynb) is the same as this lab. You can submit to the autograder the same .py file that you submitted for CS100.1x. There is no need to rename or modify the .py file.

Before submitting your assignment, please ensure that your submission follows these guidelines:

- Only use the following libraries: standard python libraries, numpy, pyspark, and test_helper (the autograder library).
- Don't include any extraneous code as the autograder will timeout if a submission takes too long.
- Only change sections of code where you see FILL IN. Changing other parts of the code, including directory paths, may cause the code to fail the autograder's tests.
- The autograder grades submissions using the same tests that are included in your notebook. Hence, please run the local tests before submitting to the autograder.

For further background on the autograder please visit this page with the information from Week 0, or the autograder FAQ. To check the status of your submission, please visit the autograder submission website.

This is the autograder only for your submission of "Lab2 - Word Count". Please **DO NOT submit other labs to this autograder.**

Lab1 from CS100.1x can be submitted here.

Choose Files No file chosen

```
Pluralize and test (1b)
----------------------
All tests passed
Apply makePlural to the base RDD(1c)
-----------------------------------
All tests passed
Pass a lambda function to map (1d)
---------------------------------
All tests passed
Length of each word (1e)
------------------------
All tests passed
Pair RDDs (1f)
--------------
All tests passed
groupByKey() approach (2a)
--------------------------
All tests passed
Use groupByKey() to obtain the counts (2b)
-----------------------------------------
All tests passed
Counting using reduceByKey (2c)
------------------------------
All tests passed
All together (2d)
-----------------
All tests passed
Unique words (3a)
-----------------
All tests passed
Mean using reduce (3b)
----------------------
All tests passed
wordCount function (4a)
-----------------------
All tests passed
Capitalization and punctuation (4b)
-----------------------------------
All tests passed
Words from lines (4d)
---------------------
All tests passed
Remove empty elements (4e)
--------------------------
All tests passed
Count the words (4f)
--------------------
All tests passed
-- 16 cases passed (100.0%) --


Your submission token ID is 1641901-0d88c48dee65e8c328e16dd6fd51f155:ip-172-31-8-27
Please include this submission token ID when you need support for your code submission.
Your anonymous student ID is d9e65bc8d252ec579ef766790c87772e. Do not post this ID on
Piazza.
```

CHECK **SAVE** *You have used 1 of 10 submissions*