OVERVIEW

This lab is consists of two notebooks: a Spark tutorial and a word count exercise.  Only the word count notebook will be graded.

The lab is due July 11, 2015 at 07:00 UTC. There is a three day grace period for late submissions until Jun 14, 2015 at 07:00 UTC. Submissions after that time will lose 20 points.

*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.*

NOTEBOOKS DOWNLOAD

The two notebooks can be found on GitHub at the following links:

* [Spark Tutorial](https://raw.githubusercontent.com/spark-mooc/mooc-setup/master/spark_tutorial_student.ipynb)
* [Word Count](https://raw.githubusercontent.com/spark-mooc/mooc-setup/master/ML_lab2_word_count_student.ipynb)

When working on these notebooks, out of respect for current (and future) students taking this course, please**do not store your solutions in publicly visible repositories** such as GitHub.  Similarly, please**do not post code snippets on Piazza**.

VIEW NOTEBOOKS

GitHub automatically renders IPython notebooks.  To view the notebooks follow these GitHub links:

* [Spark Tutorial](https://github.com/spark-mooc/mooc-setup/blob/master/spark_tutorial_student.ipynb)
* [Word Count](https://github.com/spark-mooc/mooc-setup/blob/master/ML_lab2_word_count_student.ipynb)

Note: these links should not be used to download the notebooks.  Use the raw links in the Notebooks Download section when downloading.

DETAILED INSTRUCTIONS FOR RUNNING NOTEBOOKS IN VM

1. Start the VM - To start the VM, from a DOS prompt (Windows) or Terminal (Mac/Linux), issue the command "vagrant up", while in the custom directory created for this course (you should have created this directory as part of the Week 0 segment on "[Downloading and Installing the VM Image](https://courses.edx.org/courses/BerkeleyX/CS190.1x/1T2015/courseware/9d251397874d4f0b947b606c81ccf83c/035f696c1172403f9253736f5dc2dccb/521c1b7183a74232a0385cb38f192e71)").
2. Once the VM is running, access the Jupyter web UI for running IPython notebooks by navigating your web browser to "<http://localhost:8001/>" (or "<http://127.0.0.1:8001/>").
3. Shut down any notebooks you have running, as only **ONE** notebook should run at a time.  Running notebooks have a green icon to the left of the notebook name and green text to the right of the screen that says "Running".  Shutdown running notebooks by clicking the checkbox next to the notebook and then clicking the orange "Shutdown" button.
4. Download the IPython notebooks.  **Make sure that the file extension is .ipynb**.  If the download adds an extension (e.g. ".txt"), rename the file so that the extension is just .ipynb.
   * Download the Spark Tutorial [here](https://raw.githubusercontent.com/spark-mooc/mooc-setup/master/spark_tutorial_student.ipynb).
   * Download Lab 2 Word Count [here](https://raw.githubusercontent.com/spark-mooc/mooc-setup/master/ML_lab2_word_count_student.ipynb).
5. Upload the IPython notebooks.  This process was explained in the "[Running Your First Notebook](https://courses.edx.org/courses/BerkeleyX/CS190.1x/1T2015/courseware/9d251397874d4f0b947b606c81ccf83c/035f696c1172403f9253736f5dc2dccb/41c3737c2a82464aaef6310f99c45ae6)" segment in Week 0.
6. For the **Spark Tutorial** notebook, you can simply read the text and run the cells, though feel free to change cells to see how the output changes.  If you change something and can't get a cell to run properly, re-upload the notebook to Jupyter.  This notebook **SHOULD NOT** be submitted to the autograder.
7. For the **Lab 2 Word Count** notebook, please follow the instructions in the notebook and replace <FILL IN> sections with your solutions, and submit to the autograder, following the same [guidelines](https://courses.edx.org/courses/BerkeleyX/CS190.1x/1T2015/courseware/9d251397874d4f0b947b606c81ccf83c/035f696c1172403f9253736f5dc2dccb/becab56ac25a4916ab6ac1926ae93ea2) as in previous weeks.
8. When you have submitted successfully, you can shutdown the VM by issuing the command "vagrant halt".