Assignment: 02 - Cluster Analysis

Bucket: 1 - Homework

Due Date: 2016-02-01

Grading Hidden? no

Teams? no

Description:

As data sizes will increase the single machine version of your program will not scale. Develop a version of A1 using the Hadoop Map Reduce API.

1. Group assignment, two students.
2. Provide code that can run in pseudo-distributed mode as well as on EMR.
3. Produce a graph that plots the average ticket price for each month for each airline. Use R. No other output is required.
4. Include a script that executes everything and produces the graph. For example, if you use the Unix make command, you should have two targets pseudo and cloud such that typing "make pseudo" will create a HDFS file system, start hadoop, run your job, get the output, and produce the graph. Typing "make emr" will run your code on EMR.
5. Only plot airlines with flights in 2015, limit yourself to the 10 airlines with the most flights overall.
6. Information on how to setup AWS is here (https://docs.google.com/document/d/18HkaZ0tSAH8xjDOZjPULBfAJw5H_vj33QSLemseRM2o/edit#heading=h.1yn1e27nbg83).
7. Write a one page report that documents your implementation and that describes your results. The report should be automatically constructed as part of running the project to include the plot. (Hint: use LaTeX or Markdown)
8. Submit a tar.gz file which unpacks into a directory name "LastName1_LastName2_A2". That directory should contain a README file that explains how to build and run your code. Make sure that the code is

portable. Document what it requires.

9. The reference solution builds off A1, adding 154 lines of Java code and 15 lines of R code.

Assignment Download: ()

Your Submissions

New Submission (/assignments/27/submissions/new)

Date	Status	Automatic	Teacher	Score	Link
2016-02-01 04:59:24 -0500	✓	ø / 100	100.0 / 100	100.0 / 100	View (/submissions/1982)
2016-02-01 04:45:30 -0500	?	ø / 100	ø / 100	0.0 / 100	View (/submissions/1962)
2016-02-01 04:40:08 -0500	?	ø / 100	ø / 100	0.0 / 100	View (/submissions/1954)

Course Page (http://www.ccs.neu.edu/home/ntuck/courses/2016/01/cs6240/index.html) | Piazza (https://piazza.com/class/ij4yepvnz8v3zf)

ajax-status: none