

🏠 (/) / Courses (/courses) / 6240 Parallel Data Processing with Map-Reduce (/courses/5) / 1 - Homework (/courses/5)
/ 01 - Threaded Analysis

Assignment: 01 - Threaded Analysis

Bucket: 1 - Homework

Due Date: 2016-01-23

Grading Hidden? no

Teams? no

Description:

Airfares evolve with time, one month of traffic is not sufficient to answer which airline is the cheapest. Try with two years worth of OTP data. Improve throughput of your code with parallel processing primitives.

Here's sample data: all-v1.tar (<https://s3.amazonaws.com/cs6240sp16/all-v1.tar>)

Reqs:


1. Include a README with your output and conclusions.
2. Make sure your conclusions make sense.
3. Individual assignment.
4. The input is "-p -input=DIR" where DIR is the path to a directory containing data files. All files in the directory will be processed.
5. The output is K and F, and a sequence of "C p m"s where m is the median ticket price. Restrict your output to airlines that are active in January 2015.
6. Clean up your code, document and test it.
7. Values for K and F are 128160 and 12601051. One sample airline is "AS 202.36 171.57". The reference solution is 150LOC additional/changed over A0. Processing time is under a minute.
8. Submit your assignment as a single tar.gz file which unpacks into a directory named "LastName_FirstName_A1".
9. That directory should contain a README file that explains how to build and run your assignment.

10. No .class files, .jar files, or data in your submission. Use Gradle or SBT to pull in external dependencies if necessary.

Assignment Download: ()

Your Submissions

[New Submission \(/assignments/22/submissions/new\)](/assignments/22/submissions/new)

Date	Status	Automatic	Teacher	Score	Link
2016-01-23 05:30:13 -0500		∅ / 100	100.0 / 100	100.0 / 100	View (/submissions/1507)

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

Bottlenose copyright © 2012-2015 Nat Tuck. Licensed under the GNU Affero GPL (/agpl-3.0.txt) v3 or later. Source at github (<http://www.github.com/NatTuck/bottlenose>). The development team takes no responsibility for death or serious injury that may result from use of this program.

ajax-status: none