

🏠 (/) / Courses (/courses) / 6240 Parallel Data Processing with Map-Reduce (/courses/5) / 1 - Homework (/courses/5)
/ 06 - Missed Connections (Spark)

Assignment: 06 - Missed Connections (Spark)

Bucket: 1 - Homework

Due Date: 2016-02-25

Grading Hidden? no

Teams? no

Description:

Compute the percentage of missed connections for all two-hop paths. Use Spark.



- Group assignment, two students.
- A connection is any pair of flight F and G of the same carrier such as `F.Destination = G.Origin` and the scheduled departure of G is ≤ 6 hours and ≥ 30 minutes after the scheduled arrival of F.
- A connection is missed when the actual arrival of F < 30 minutes before the actual departure of G.
- Optimize your code for performance.
- Output the number of missed connections per airline, per year.
- In your report, compare the performance of your Map-Reduce and Spark implementations.
 - Test running locally.
 - Test running on a 4-worker cluster on EMR.

Submit early. "The server was overloaded from 11pm-midnight" is not an excuse for a late submission.

Assignment Download: ()

Your Submissions

New Submission (/assignments/53/submissions/new)

Date	Status	Automatic	Teacher	Score	Link
2016-03-06 13:28:09 -0500		∅ / 100	100.0 / 100	100.0 / 100	View (/submissions/4999)
2016-03-01 00:44:58 -0500		∅ / 100	∅ / 100	0.0 / 100	View (/submissions/4466)

Course Page (<http://www.ccs.neu.edu/home/ntuck/courses/2016/01/cs6240/index.html>) | Piazza (<https://piazza.com/class/ij4yepvz8v3zf>)

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