

Python Equi-join MapReduce Homework

Mapper Input:

[table_identifier, join_key, sequence of other columns]

- ["order", "32", "130057", "O", "208660.75", "1995-07-16", "2-HIGH", "Clerk#000000616", "0", "ise blithely bold, regular requests. quickly unusual dep"]
- ...
- ["line_item", "32", "82704", "7721", "1", "28", "47227.60", "0.05", "0.08", "N", "O", "1995-10-23", "1995-08-27", "1995-10-26", "TAKE BACK RETURN", "TRUCK", "sleep quickly. req"]
- ...

Reducer Output:

[sequence of the columns of ORDER, sequence of the columns of LINEITEM]

- ["order", "32", "130057", "O", "208660.75", "1995-07-16", "2-HIGH", "Clerk#000000616", "0", "ise blithely bold, regular requests. quickly unusual dep", "line_item", "32", "82704", "7721", "1", "28", "47227.60", "0.05", "0.08", "N", "O", "1995-10-23", "1995-08-27", "1995-10-26", "TAKE BACK RETURN", "TRUCK", "sleep quickly. req"]

>> python **join.py** records.json

NOTE:

Zip **join.py** & a **report** about what you learned, outputs, and command line you executed. Then, upload zip file (named your student id like assignment1_2012104087-ORDER.zip)

Plagiarism:

Your python file will be checked by each other students' file to prevent 'plagiarism'. FYI, a professor has complete codes that people submitted before. If you copy and look at the other python file, you will get the lowest grade.

Copyright:

This material was provided by Prof. Bill Howe at University of Washington for his Coursera course "Introduction to Data Science.". Prof. Hae Joon Lee in Kyung-Hee University got permission to use this assignment for our course "Big Data Programming".