

Assignment 1

Write a mapper.py and reducer.py for a MapReduce job to compute the average of personal incomes for each district.

Each line in the input data file should contain tuples of (person id, district id, personal income) as follows:

10021,2,120000

10023,3,200000

14500,2,320000

10021,1,500000

.....

.....

Create your own data for test run. Submit the output screen and the results from the test run on the quickstart VM.