Problem:

Write a mapper and reducer program to join two databases?

Consider Input data

joinA.txt

able,991

about,11

burger,15

actor,22

about,13

joinB.txt

Jan-01 able,5

Feb-02 about,3

Mar-03 about,8

Apr-04 able,13

Feb-22 actor,3

Feb-23 burger,5

Mar-08 burger,2

Dec-15 able,100

Finally, the output should look like : join_output.txt

('able', 'Dec-15', '100', '991')

('able', 'Apr-04', '13', '991')

('able', 'Jan-01', '5', '991')

('about', 'Mar-03', '8', '13')

('about', 'Feb-02', '3', '13')

('about', 'Mar-03', '8', '11')

('about', 'Feb-02', '3', '11')

('actor', 'Feb-22', '3', '22')

('burger', 'Feb-23', '5', '15')

('burger', 'Mar-08', '2', '15')

General Instruction:

1) Create directory for input or choose address according to your choice.
2) Put input data files into hdfs
3) Create mapper and reducer program
4) Call Hadoop
5) Download result
   hdfs dfs -getmerge /user/hadoop/output/* join_output.txt