scp at3857@access.cims.nyu.edu:/home/at3857/dataset3.csv /home/at3857

hdfs dfs -put dataset3.csv /user/at3857/project2

hdfs dfs -chmod a+x /user/at3857/project2

Input file:

/user/at3857/project2/dataset3.csv

Output file:

/user/at3857/project2/output3

Mapper file:

/user/at3857/project/pragyamapper.py

hadoop jar /opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop-mapreduce/hadoop-streaming.jar -D mapreduce.job.reduces=1 -files hdfs://dumbo/user/at3857/project/ pragyamapper.py -mapper "python pragyamapper.py" -input /user/at3857/project2/dataset3.csv -output /user/at3857/project2/output3

cat ./dataset3.csv | python ./pragyamapper.py

create external table dataset3new (year int,industry string, code string, cipname string, cipcode string)

row format delimited fields terminated by '|'

location '/user/at3857/project2/output3/';