**Big Data Homework-2**

Task:-

1. Submit a copy of this modified program and a screen shot of the results of the program’s execution as the output of your assignment.

PROGRAM:-

from mrjob.job import MRJob

import re

WORD\_RE = re.compile(r"[\w']+")

class MRWordCount(MRJob):

def mapper(self, \_, line):

for word in WORD\_RE.findall(line):

if word[0] >= "a" and word[0] <= "n":

yield "a\_to\_n", 1

else:

yield "other", 1

def combiner(self, word, counts):

yield word, sum(counts)

def reducer(self, word, counts):

yield word, sum(counts)

if \_\_name\_\_ == '\_\_main\_\_':

MRWordCount.run()

OUTPUT:-



1. Submit a copy of this modified program and a screen shot of the results of the program’s execution as the output of your assignment.

PROGRAM:-

from mrjob.job import MRJob

class MRSalaries(MRJob):

def mapper(self, \_, line):

(name,jobTitle,agencyID,agency,hireDate,annualSalary,grossPay) = line.split('\t')

pay = float(annualSalary)

if pay >= 0.00 and pay <= 49999.99 :

yield "Low", 1

if pay >= 50000.00 and pay <= 99999.99 :

yield "Medium", 1

if pay >= 100000.00 :

yield "High", 1

def combiner(self, pay, counts):

yield pay, sum(counts)

def reducer(self, pay, counts):

yield pay, sum(counts)

if \_\_name\_\_ == '\_\_main\_\_':

MRSalaries.run()

OUTPUT:-



1. Submit a copy of the program for u.data and a screen shot of the results of the program’s execution (only 10 lines or so of the result) as the output of your assignment.

PROGRAM:-

from mrjob.job import MRJob

class MRMovieCount(MRJob):

def mapper(self, \_, line):

(USERID,MOVIEID,RATING,TIMESTAMP) = line.split(',')

yield USERID, 1

def combiner(self, USERID, counts):

yield USERID, sum(counts)

def reducer(self, USERID, counts):

yield USERID, sum(counts)

if \_\_name\_\_ == '\_\_main\_\_':

MRMovieCount.run()

OUTPUT:-

