**TERMWORK4**

**Term4data.csv**

1,M,25000,2,Agree

2,F,50000,1,DisAgree

3,M,75000,0,Neutral

4,F,80000,2,Agree

5,F,10000,1,DisAgree

6,F,20000,3,Neutral

7,M,17000,0,DisAgree

8,F,15000,0,DisAgree

9,M,60000,1,Agree

10,F,45000,1,Agree

11,F,46000,3,DisAgree

12,F,50000,3,Neutral

**Mapper4.py**

#!/usr/bin/env python

import sys

for line in sys.stdin:

line = line.strip()

line = line.split(",")

if len(line) >=2:

pid = line[0]

opinion = line[4]

print '%s\t%s' % (pid, opinion)

**Reducer4.py**

#!/usr/bin/env python

import sys

opiniondic={}

count=0

for line in sys.stdin:

line = line.strip()

pid, opinion = line.split('\t')

if opinion in opiniondic:

opiniondic[opinion].append(count+1)

else:

opiniondic[opinion] = []

opiniondic[opinion].append(count+1)

for op in opiniondic.keys():

count=len(opiniondic[op])

print '%s\t%s'% (op,count)

**TERMWORK5**

**Term5data.csv write in csv(,)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| E001 | Sunita | Accounts | 15000 |  |
| E002 | Harsh | IT | 50000 |  |
| E003 | Ragini | IT | 75000 |  |
| E004 | Mithun | Accounts | 20000 |  |
| E005 | Pruthavi | Marketing | 45000 |  |
| E006 | Anjali | IT | 70000 |  |
| E007 | Kunal | Marketing | 60000 |  |
| E008 | Mitali | Accounts | 55000 |  |
| E009 | Roopa | IT | 70000 |  |
| E010 | Deepti | Accounts | 30000 |  |
| E011 | Janavi | Marketing | 25000 |  |
| E012 | Lata | Accounts | 30000 |  |
| E013 | Brijmohan | Marketing | 45000 |  |
| E014 | Nina | Accounts | 50000 |  |
| E015 | Pallavi | Marketing | 25000 |  |

**Mapper5.py**

#!/usr/bin/env python

import sys

# input comes from STDIN (standard input)

for line in sys.stdin:

line = line.strip()

line = line.split(",")

if len(line) >=2:

dept = line[2]

sal = line[3]

print '%s\t%s' % (dept, sal)

**Reducer5.py**

#!/usr/bin/env python

import sys

deptdic={}

for line in sys.stdin:

line = line.strip()

dept,sal = line.split('\t')

if dept in deptdic:

deptdic[dept].append(int(sal))

else:

deptdic[dept] = []

deptdic[dept].append(int(sal))

for dept in deptdic.keys():

sum\_sal = sum(deptdic[dept])

print '%s\t%s'% (dept,sum\_sal)