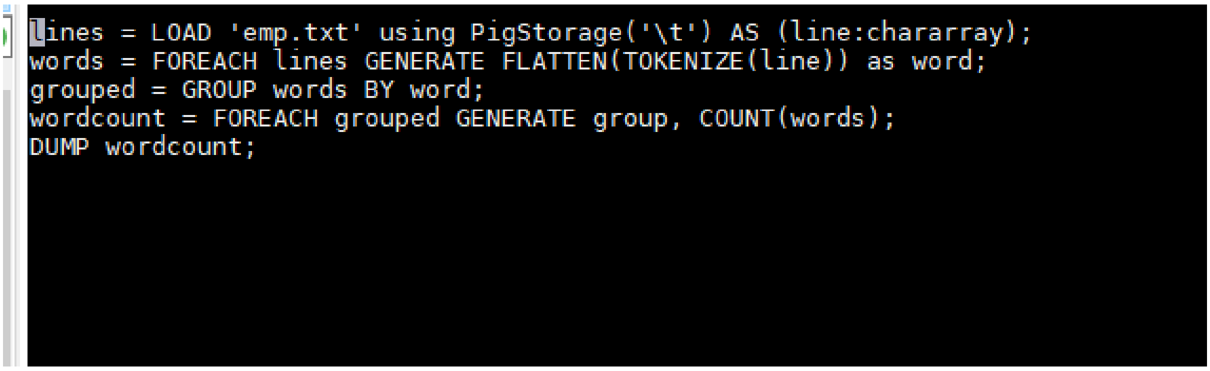
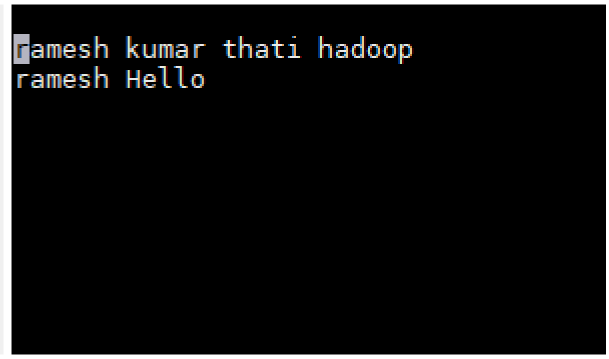
Write a program to implement wordcount using Pig.

Code for WordCount



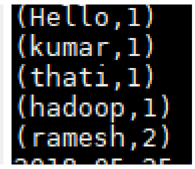
```
lines = LOAD 'emp.txt' using PigStorage('\t') AS (line:chararray);
words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;
grouped = GROUP words BY word;
wordcount = FOREACH grouped GENERATE group, COUNT(words);
DUMP wordcount;
```

Input:



```
ramesh kumar thati hadoop
ramesh Hello
```

Output:

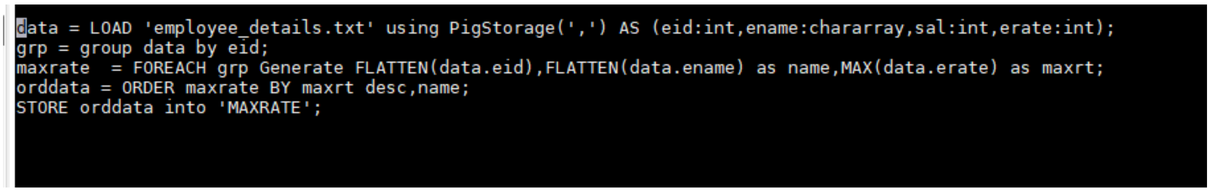


```
(Hello,1)
(kumar,1)
(thati,1)
(hadoop,1)
(ramesh,2)
```

Task-2

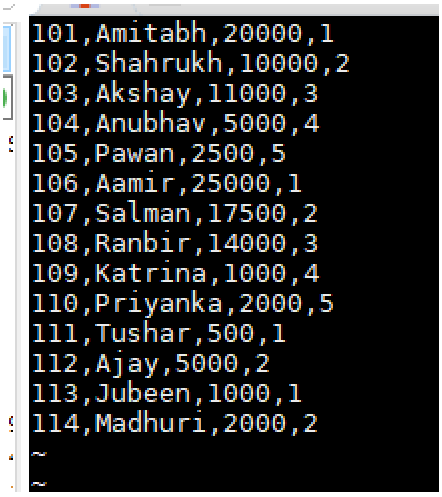
a. Top 5 employees (employee id and employee name) with highest rating. (In case two employees have same rating, employee with name coming first in dictionary should get preference.

   Source Code:



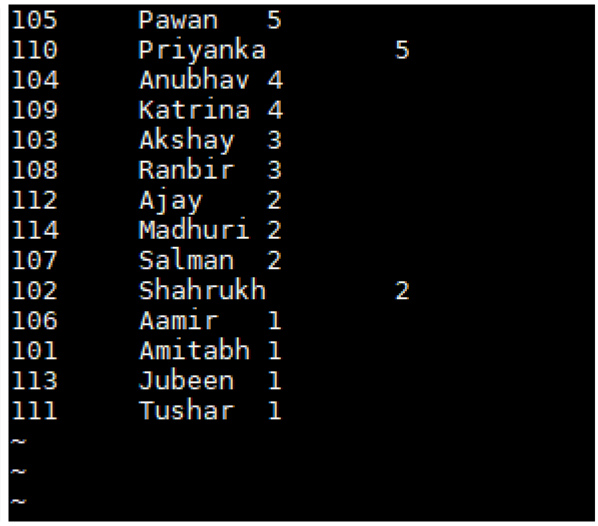
```
data = LOAD 'employee_details.txt' using PigStorage(',') AS (eid:int,ename:chararray,sal:int,erate:int);
grp = group data by eid;
maxrate  = FOREACH grp Generate FLATTEN(data.eid),FLATTEN(data.ename) as name,MAX(data.erate) as maxrt;
orddata = ORDER maxrate BY maxrt desc,name;
STORE orddata into 'MAXRATE';
```

Input



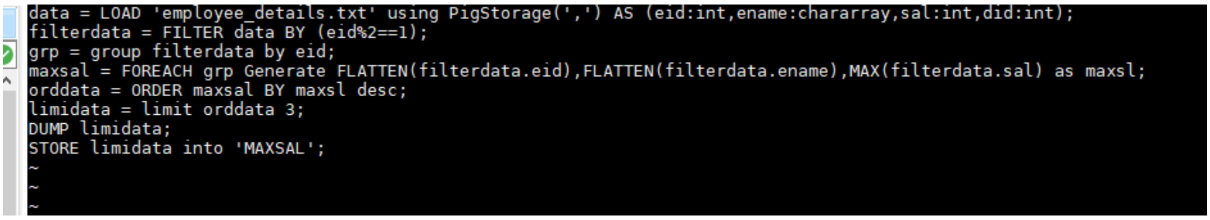
```
101,Amitabh,20000,1
102,Shahrukh,10000,2
103,Akshay,11000,3
104,Anubhav,5000,4
105,Pawan,2500,5
106,Aamir,25000,1
107,Salman,17500,2
108,Ranbir,14000,3
109,Katrina,1000,4
110,Priyanka,2000,5
111,Tushar,500,1
112,Ajay,5000,2
113,Jubeen,1000,1
114,Madhuri,2000,2
~
~
```

Output:



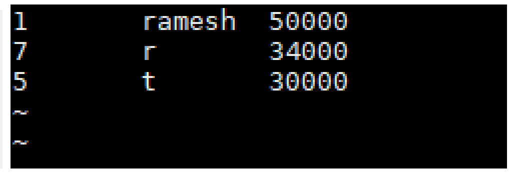
```
105     Pawan   5
110     Priyanka        5
104     Anubhav 4
109     Katrina 4
103     Akshay  3
108     Ranbir  3
112     Ajay    2
114     Madhuri 2
107     Salman  2
102     Shahrukh        2
106     Aamir   1
101     Amitabh 1
113     Jubeen  1
111     Tushar  1
~
~
~
```

(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id is an odd number. (In case two employees have same salary, employee with name coming first in dictionary should get preference



```
data = LOAD 'employee_details.txt' using PigStorage(',') AS (eid:int,ename:chararray,sal:int,did:int);
filterdata = FILTER data BY (eid%2==1);
grp = group filterdata by eid;
maxsal = FOREACH grp Generate FLATTEN(filterdata.eid),FLATTEN(filterdata.ename),MAX(filterdata.sal) as maxsl;
orddata = ORDER maxsal BY maxsl desc;
limidata = limit orddata 3;
DUMP limidata;
STORE limidata into 'MAXSAL';
~
~
~
```

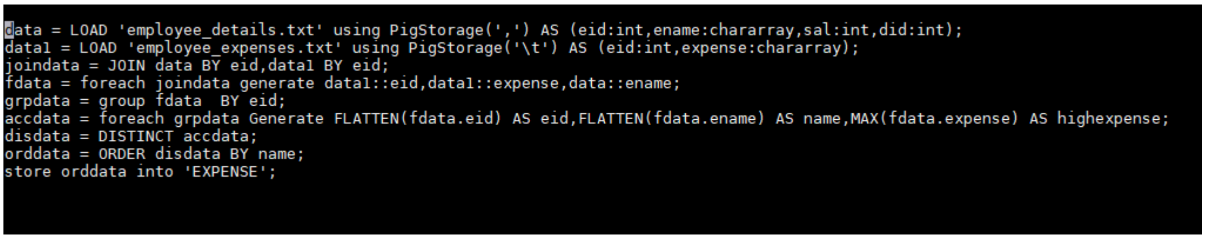
Output



```
1       ramesh  50000
7       r       34000
5       t       30000
~
~
```

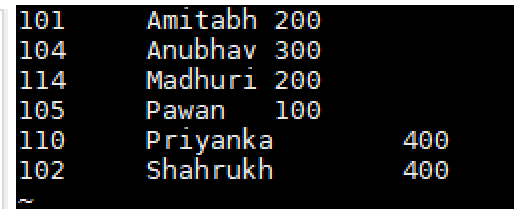
(c) Employee (employee id and employee name) with maximum expense (In case two employees have same expense, employee with name coming first in dictionary should get preference).

Source Code:



```
data = LOAD 'employee_details.txt' using PigStorage(',') AS (eid:int,ename:chararray,sal:int,did:int);
data1 = LOAD 'employee_expenses.txt' using PigStorage('\t') AS (eid:int,expense:chararray);
joindata = JOIN data BY eid,data1 BY eid;
fdata = foreach joindata generate data1::eid,data1::expense,data::ename;
grpdata = group fdata  BY eid;
accdata = foreach grpdata Generate FLATTEN(fdata.eid) AS eid,FLATTEN(fdata.ename) AS name,MAX(fdata.expense) AS highexpense;
disdata = DISTINCT accdata;
orddata = ORDER disdata BY name;
store orddata into 'EXPENSE';
```

Output:

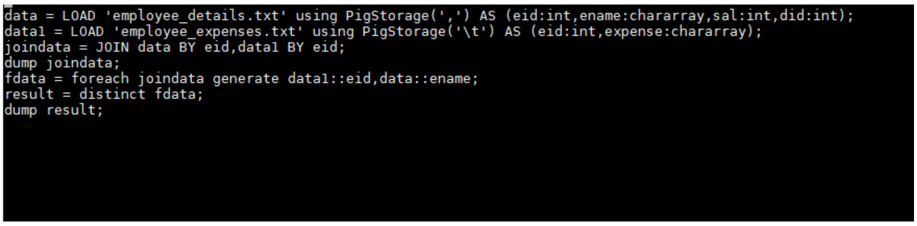


```
101     Amitabh 200
104     Anubhav 300
114     Madhuri 200
105     Pawan   100
110     Priyanka        400
102     Shahrukh        400
~
```

(d) List of employees (employee id and employee name) having entries in employee_expenses

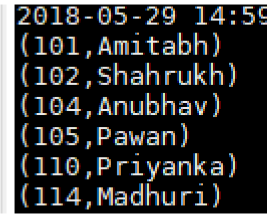
file.

Source Code:



```
data = LOAD 'employee_details.txt' using PigStorage(',') AS (eid:int,ename:chararray,sal:int,did:int);
data1 = LOAD 'employee_expenses.txt' using PigStorage('\t') AS (eid:int,expense:chararray);
joindata = JOIN data BY eid,data1 BY eid;
dump joindata;
fdata = foreach joindata generate data1::eid,data::ename;
result = distinct fdata;
dump result;
```

OutPut-file

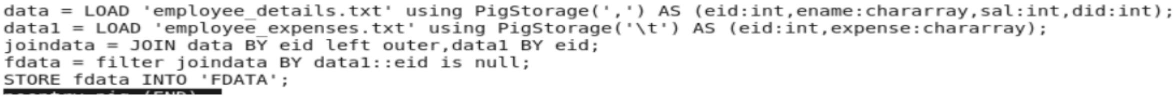


```
2018-05-29 14:59
(101,Amitabh)
(102,Shahrukh)
(104,Anubhav)
(105,Pawan)
(110,Priyanka)
(114,Madhuri)
```

(e) List of employees (employee id and employee name) having no entry in employee_expenses

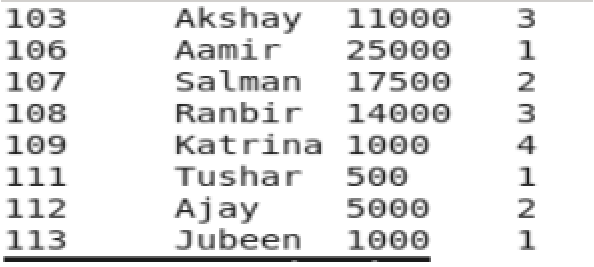
file.

Source File



```
data = LOAD 'employee_details.txt' using PigStorage(',') AS (eid:int,ename:chararray,sal:int,did:int);
data1 = LOAD 'employee_expenses.txt' using PigStorage('\t') AS (eid:int,expense:chararray);
joindata = JOIN data BY eid left outer,data1 BY eid;
fdata = filter joindata BY data1::eid is null;
STORE fdata INTO 'FDATA';
```

Output



```
103       Akshay  11000    3
106       Aamir   25000    1
107       Salman  17500    2
108       Ranbir  14000    3
109       Katrina 1000     4
111       Tushar  500      1
112       Ajay    5000     2
113       Jubeen  1000     1
```