Question (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)

Command:

A = LOAD '/home/acadgild/hadoop/employee\_details.txt' USING PigStorage(',') as (id:int, name:chararray, salary:int, rating:int);

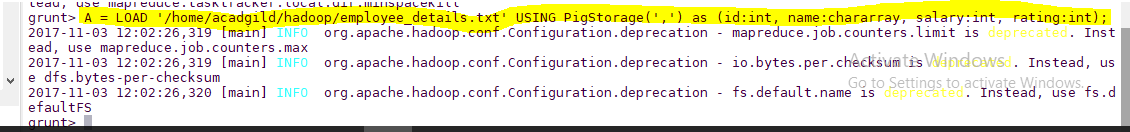
grpd = GROUP A BY rating;

top5 = FOREACH grpd {sorted = ORDER A by name ASC; top1 = limit sorted 1; generate FLATTEN(top1);};

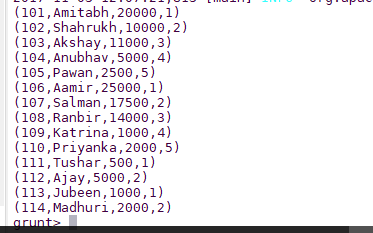
grpd2 = FOREACH top5 GENERATE emp\_id, name;

dump grpd2;

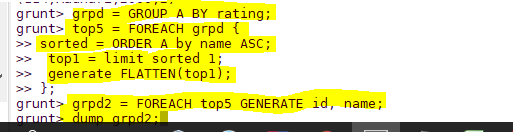
Input Command Screen-Shot



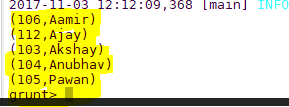
Output Screen-Shot



Input Command Screen-shot



Output Command Screen-shot



Question (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)

Command

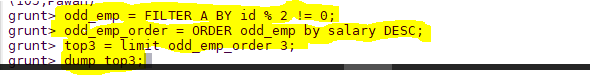
odd\_emp = FILTER A BY id % 2 != 0;

odd\_emp\_order = ORDER odd\_emp by salary DESC;

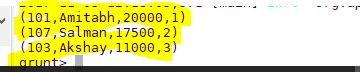
top3 = limit odd\_emp\_order 3;

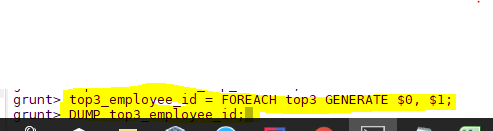
top3\_employee\_id = FOREACH top3 GENERATE $0, $1;

Input Screen-shot



Output Screen-shot







Question (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)

Command

A = LOAD '/home/acadgild/hadoop/employee\_details.txt' USING PigStorage(',') as (id:int, name:chararray, salary:int, rating:int);

B = LOAD '/home/acadgild/hadoop/employee\_expenses.txt' USING PigStorage('\t') as (id:int, expenses:int);

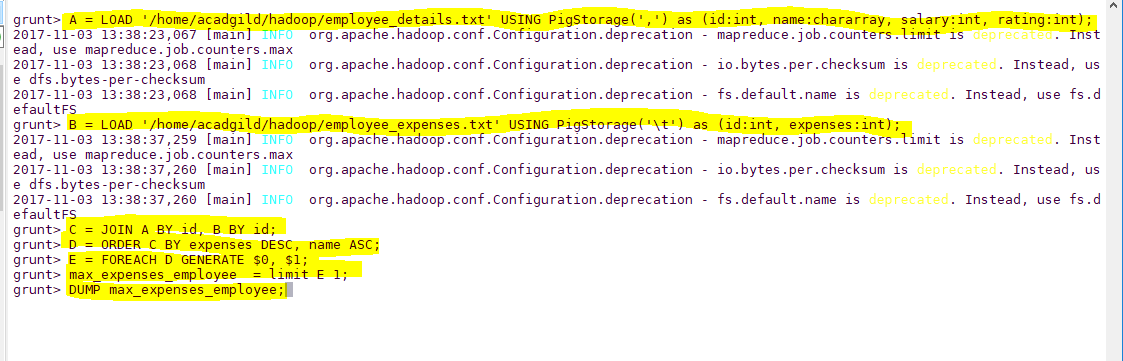
C = JOIN A BY id, B BY id;

D = ORDER C BY expenses DESC, name ASC;

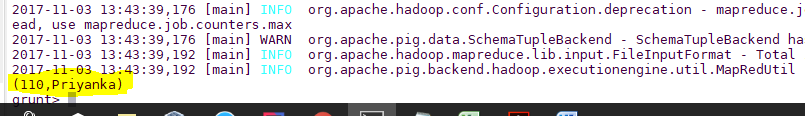
E = FOREACH D GENERATE $0, $1;

max\_expenses\_employee = limit E 1;

Input Screen-shot



Output Scree-Shot



Question(d) List of employees (employee id and employee name) having entries in employee\_expenses

file.

A = LOAD '/home/acadgild/hadoop/employee\_details.txt' USING PigStorage(',') as (id:int, name:chararray, salary:int, rating:int);

B = LOAD '/home/acadgild/hadoop/employee\_expenses.txt' USING PigStorage('\t') as (id:int, expenses:int);

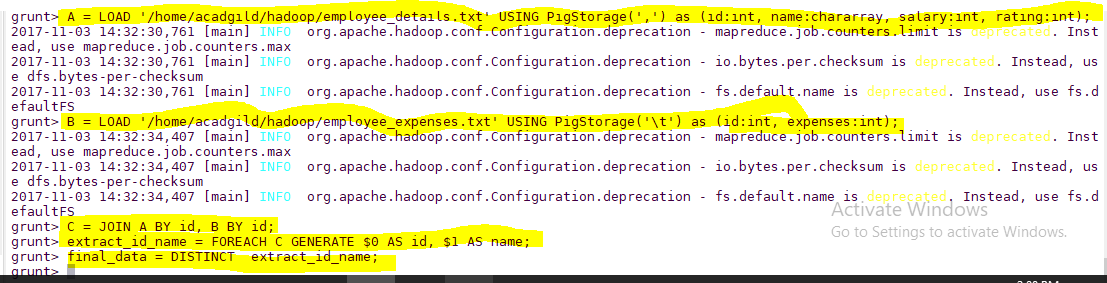
C = JOIN A BY id, B BY id;

extract\_id\_name = FOREACH C GENERATE $0 AS id, AS name;

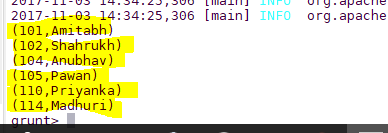
final\_data = DISTINCT extract\_id\_name;

DUMP final\_data;

Input Screen-Shot



Output Screen-Shot



Question(e) List of employees (employee id and employee name) having no entry in employee\_expenses file.

A = LOAD '/home/acadgild/hadoop/employee\_details.txt' USING PigStorage(',') as (id:int, name:chararray, salary:int, rating:int);

B = LOAD '/home/acadgild/hadoop/employee\_expenses.txt' USING PigStorage('\t') as (id:int, expenses:int);

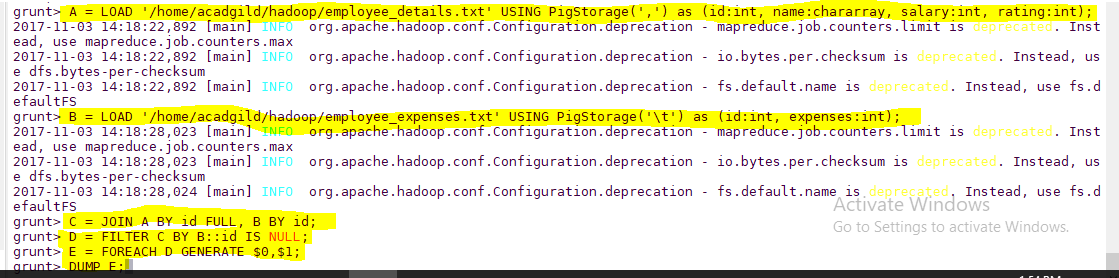
C = JOIN A BY id FULL, B BY id;

D = FILTER C BY B::id IS NULL;

E = FOREACH D GENERATE $0,$1;

DUMP E;

Input Screen-shot



Output Screen-shot

