\*\*We have employee\_details and employee\_expenses files.

>pig -x local

grunt>emp\_dtls = LOAD ' employee\_details.txt' USING PigStorage(',') AS (emp\_id:int, emp\_name:chararray, salary:long, rating:int);

grunt>emp\_expenses = LOAD ' employee\_expenses.txt' USING PigStorage(',') AS (emp\_id:int, expenses:int);

(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)

grunt>emp\_with\_high\_rating = ORDER emp\_dtls BY rating DESC;

grunt>Result1 = LIMIT emp\_with\_high\_rating 5;

> result = FOREACH Result1 GENERATE emp\_id AS ID, emp\_name AS NAME;

>DUMP result;

(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)

grunt>emp\_with\_high\_salary = ORDER emp\_dtls BY salary DESC;

grunt>emp\_with\_oddid = FILTER ' emp\_with\_high\_salary' BY emp\_id%2!=0;

grunt>Result2 = LIMIT emp\_with\_oddid 3;

>result = FOREACH Result2 GENERATE emp\_id AS ID, emp\_name AS NAME;

>DUMP result;

(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)

>joined\_data = join emp\_dtls by emp\_id , emp\_expenses by emp\_id;

>emp\_with\_high\_expense = ORDER joined\_data BY emp\_expenses::expenses DESC;

>result = LIMIT emp\_with\_high\_expense 1;

>result1 = FOREACH result GENERATE emp\_dtls::emp\_id AS id, emp\_dtls::emp\_name AS name;

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

>right\_outer\_joined\_data = join emp\_dtls by emp\_id right outer, emp\_expenses by emp\_id;

>result = FOREACH right\_outer\_joined\_data GENERATE emp\_dtls::emp\_id AS ID, emp\_dtls::emp\_name AS NAME;

>result\_1 = DISTINCT result;

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

>cogrp\_data = COGROUP emp\_dtls BY emp\_id,emp\_expenses BY emp\_id;

>flir\_data = FILTER cogrp\_data BY IsEmpty(emp\_expenses);

>result = FOREACH fltr\_data GENERATE group, emp\_dtls.emp\_name AS NAME:

>DUMP result;