**Assignment 5.1**

**Problem Statement:**

We have employee\_details and employee\_expenses files. Use local mode while running Pig and write Pig Latin script to get below results:

<https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee_details.txt>

employee id ,employee name, salary,rating

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

<https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee_expenses.txt>

employee id ,employee expense

101 200

102 100

110 400

114 200

119 200

105 100

101 100

104 300

102 400

(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)

step 1: A = LOAD 'employee\_details.txt' USING PigStorage(',') AS (id:int, name:chararray, salary:int, ratings:int);

step 2: ARating orderRating = ORDER A by ratings desc,name ASC;

step 3: orderRating = ORDER ARating by ratings ASC,name ASC;

step 4: limitedorderRating = LIMIT orderRating 5

step 5: ARating = FOREACH limitedorderRating GENERATE id,name, ratings;

step 6: dump ARating;

(105,Pawan,5)

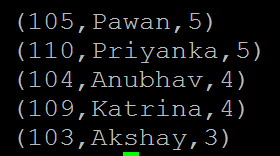
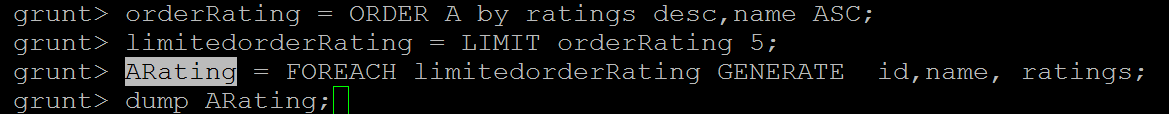
(110,Priyanka,5)

(104,Anubhav,4)

(109,Katrina,4)

(103,Akshay,3)





(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)

ASalary = FOREACH A GENERATE id, name, salary;

oddNumberSalary = FILTER ASalary BY (id %2 !=0) ;

orderSalary = ORDER oddNumberSalary by salary Desc,name ASC;

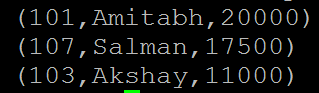
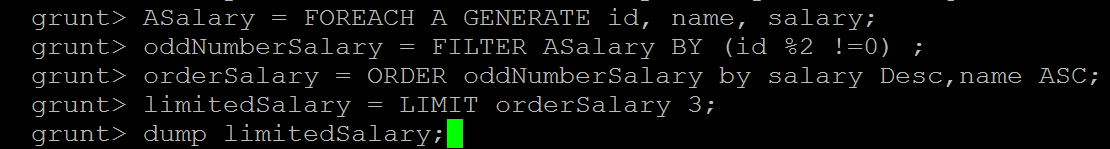
limitedSalary = LIMIT orderSalary 3;

dump limitedSalary;

(101,Amitabh,20000)

(107,Salman,17500)

(103,Akshay,11000)



(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)

expense = LOAD 'employee\_expenses.txt'

USING PigStorage('\t') AS (id:int, expenses:int);

joined\_table = join A by id, expense by id;

joined\_table\_expense = FOREACH joined\_table GENERATE $0 as id,$1 as name,$5 as expenses;

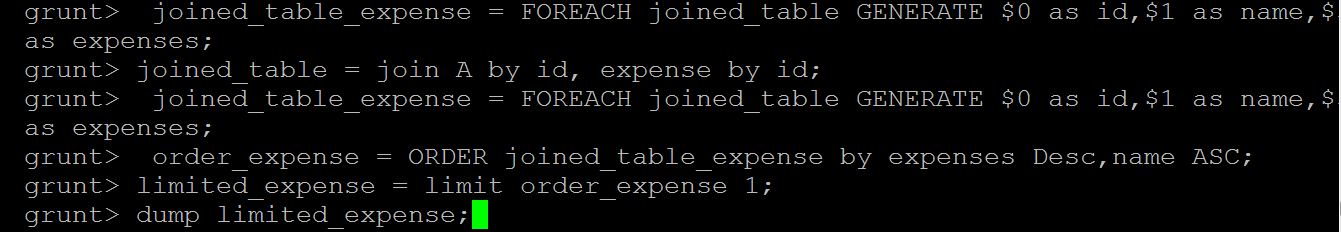
order\_expense = ORDER joined\_table\_expense by expenses Desc,name ASC;

limited\_expense = limit order\_expense 1;

dump limited\_expense;

(110,Priyanka,400)







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

emp\_expense = join A by id, expense by id;

empInExpense = foreach emp\_expense generate $0,$1;

distinctEmpInExpense = distinct empInExpense;

dump distinctEmpInExpense;

(101,Amitabh)

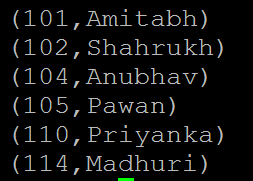
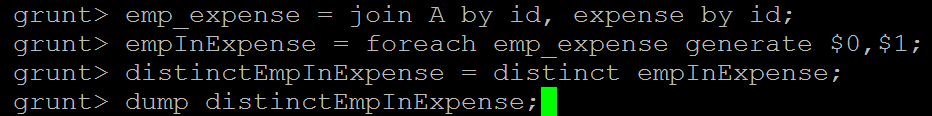
(102,Shahrukh)

(104,Anubhav)

(105,Pawan)

(110,Priyanka)

(114,Madhuri)



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

emp\_expense\_full = join A by id FULL OUTER, expense by id;

empNotInExpense\_filter = filter emp\_expense\_full by $4 is null and $5 is null;

empNotInExpense = foreach empNotInExpense\_filter generate $0,$1;

distnctEmpNotInExpense = distinct empNotInExpense;

dump distnctEmpNotInExpense;

(103,Akshay)

(106,Aamir)

(107,Salman)

(108,Ranbir)

(109,Katrina)

(111,Tushar)

(112,Ajay)

(113,Jubeen)

