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.t

xt

https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee\_expense

s.txt

(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)

**Answers**

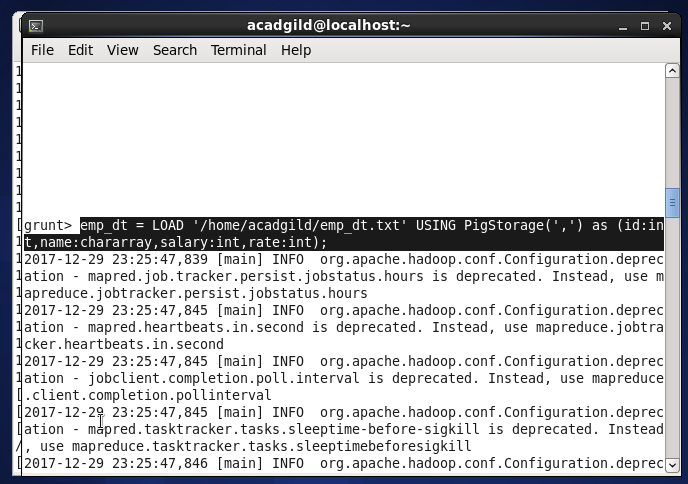
1. Emp\_dt = LOAD ‘/home/acadgild/emp\_dt.txt’ USING PigStorage(‘,’) as (id:int,name:chararray,salary:int,rating:int);

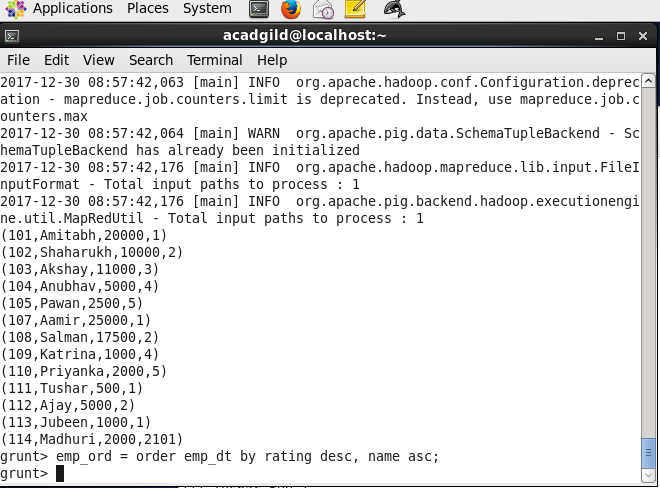
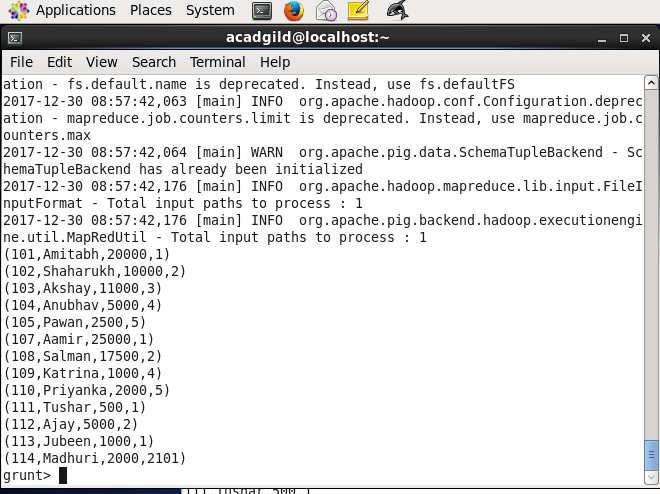
emp\_ord = order emp\_dt by rating desc, name asc;

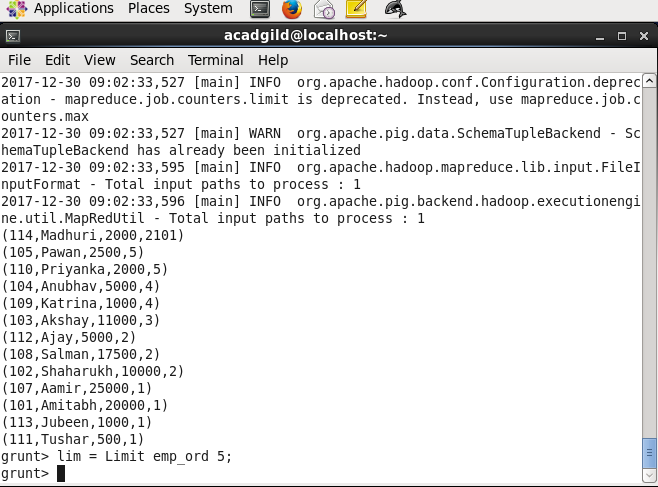
lim = limit emp\_order 5;

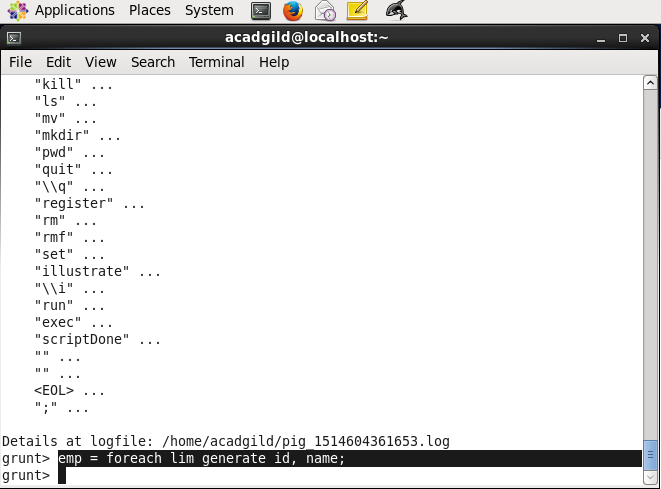
emp = foreach lim generate id, name;

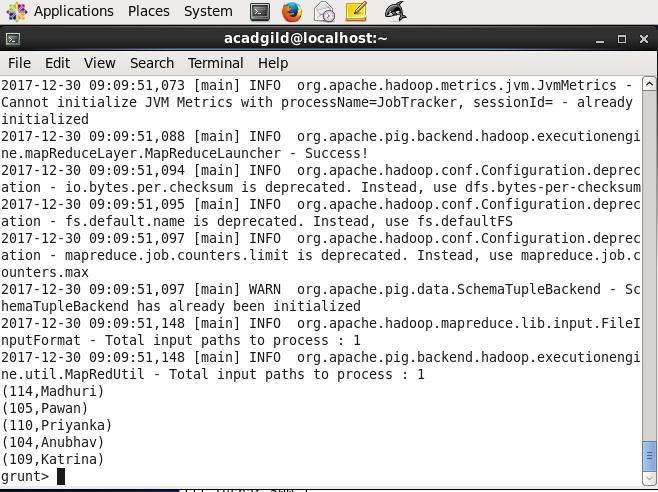
dump emp;

****

****

****

****

****

(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)

Answer :

Emp\_det = LOAD ‘/home/acadgild/emp\_dt.txt’ USING PigStorage(‘,’) as (id:int,name:chararray,salary:int,rating:int);

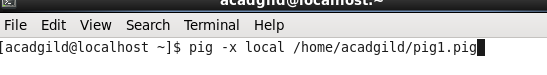
emp\_fil = filter emp\_det by (id % 2) != 0;

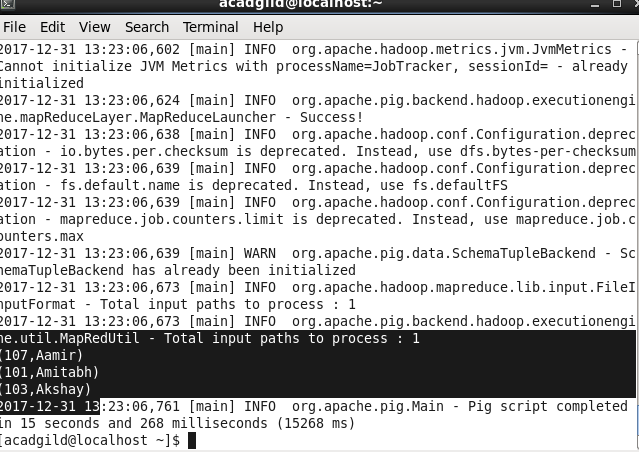
emp1\_order = order emp\_fil by salary desc;

lim1 = limit emp1\_order 3;

emp1 = foreach lim1 generate id, name;

dump emp1;





(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)

Answer :

Emp\_det = LOAD ‘/home/acadgild/emp\_dt.txt’ USING PigStorage(‘,’) as (id:int,name:chararray,salary:int,rating:int);

Emp\_exp = LOAD ‘/home/acadgild/emp\_exp.txt’ USING PigStorage(‘,’) as (id:int, expenses:int);

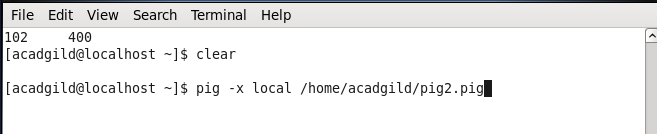
emp\_join1 = join emp\_det by id, emp\_exp by id;

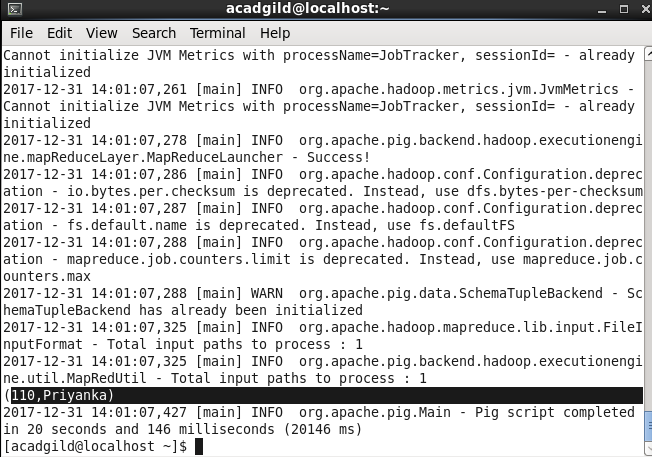
emp2\_order = order emp\_join1 by expenses desc, name asc;

lim2 = limit emp2\_order 1;

emp2 = foreach lim2 generate $0,$1;

dump emp2;





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

file.

Emp\_det = LOAD ‘/home/acadgild/emp\_dt.txt’ USING PigStorage(‘,’) as (id:int,name:chararray,salary:int,rating:int);

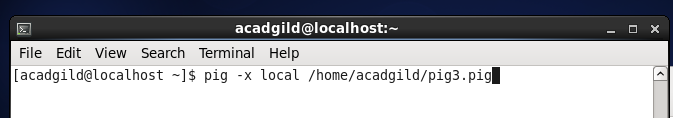
Emp\_exp = LOAD ‘/home/acadgild/emp\_exp.txt’ USING PigStorage(‘,’) as (id:int, expenses:int);

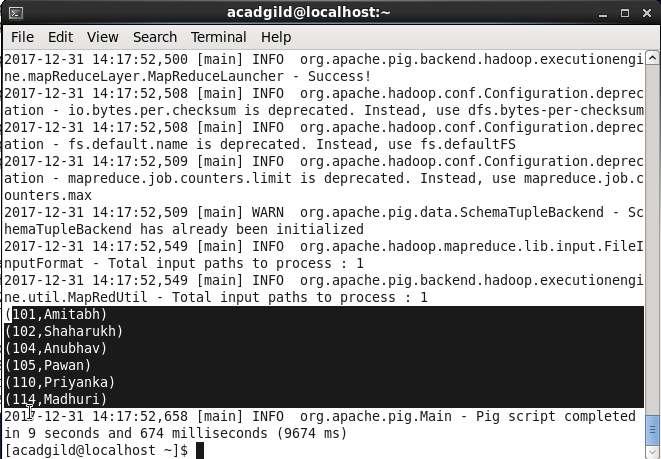
emp\_join2 = join emp\_det by id, emp\_exp by id;

emp3 = foreach emp\_join2 generate $0,$1;

distnct = distinct emp3;

dump distnct;





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

file.

emp\_join3 = join emp\_det by id left outer, emp\_exp by id;

emp4\_fil = filter emp\_join3 by $4 is null and $5 is null;

emp4 = foreach emp4\_fil generate $0,$1;

distnct1 = distinct emp4;

dump distnct1;

