**1. Introduction**

This assignment will help you to consolidate the concepts learnt in the session.

**2. 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.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)

(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)

(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)

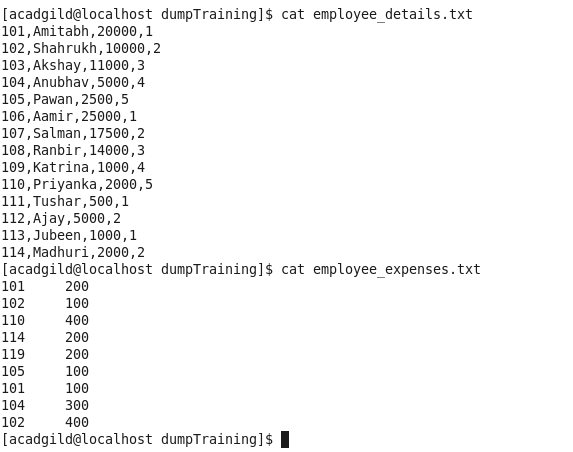
(d) List of employees (employee id and employee name) having entries in employee\_expenses

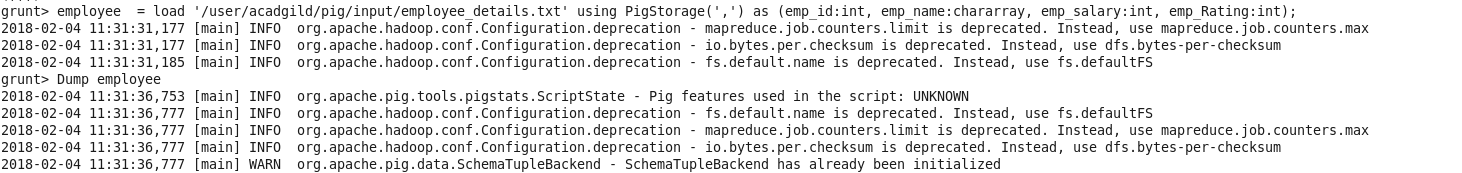
file.

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

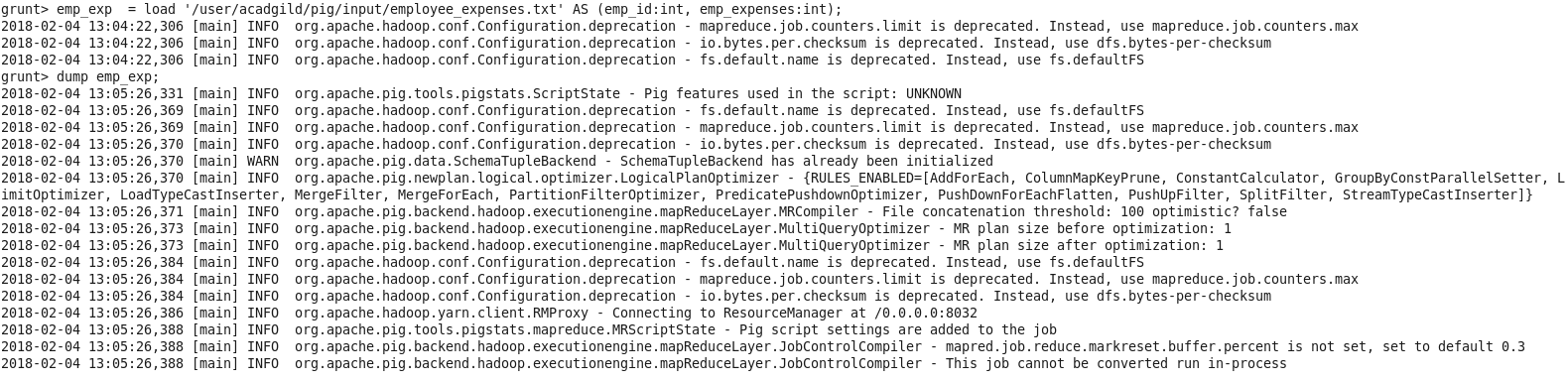
file.

**3. Output**









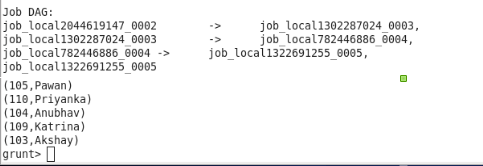


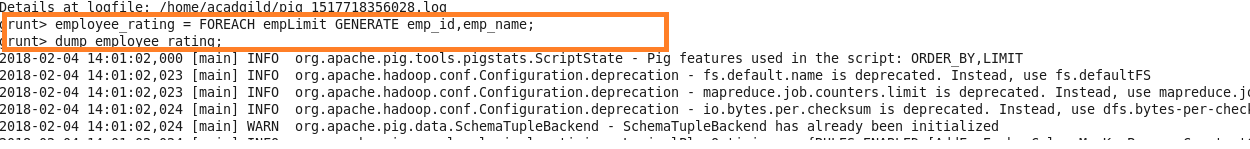
(a) Top 5 employees (employee id and employee name) with highest rating. (In case two

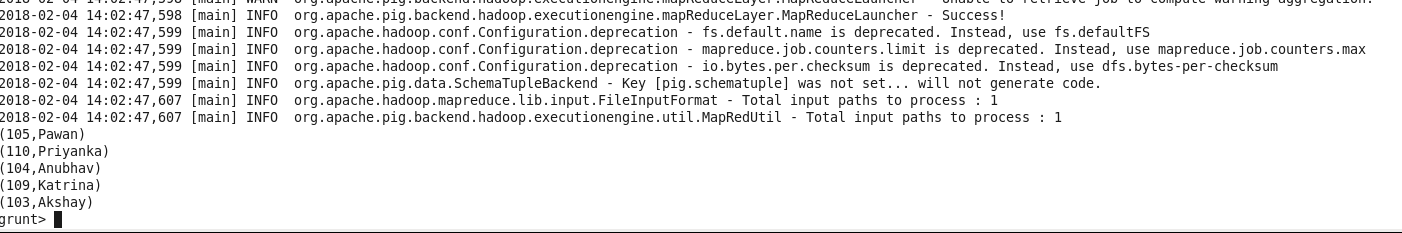
employees have same rating, employee with name coming first in dictionary should get

preference)





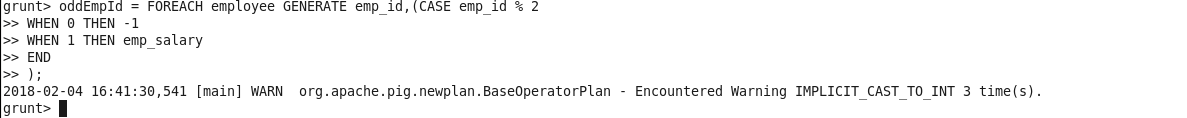




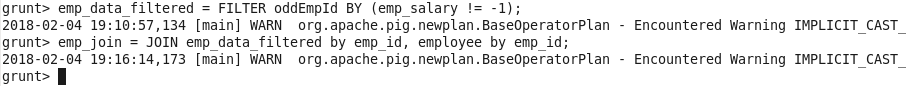
(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id

is an odd number. (In case two employees have same salary, employee with name coming first

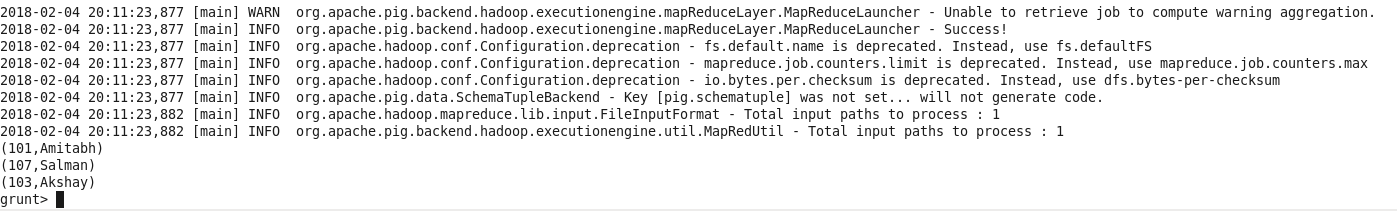
in dictionary should get preference)







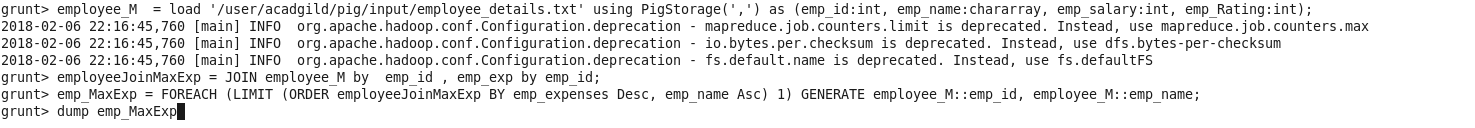


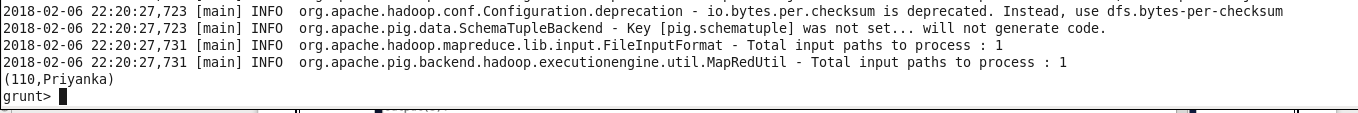


(c) Employee (employee id and employee name) with maximum expense (In case two

employees have same expense, employee with name coming first in dictionary should get

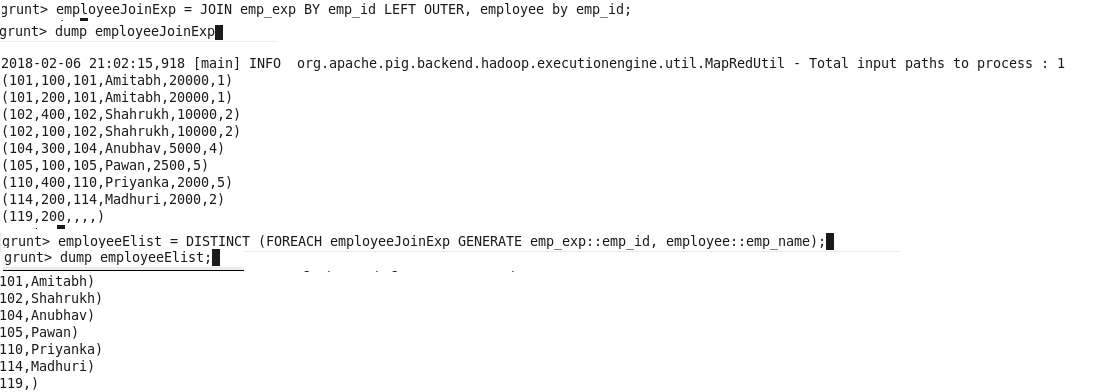
preference)



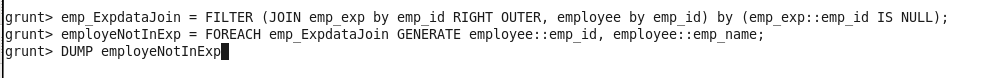


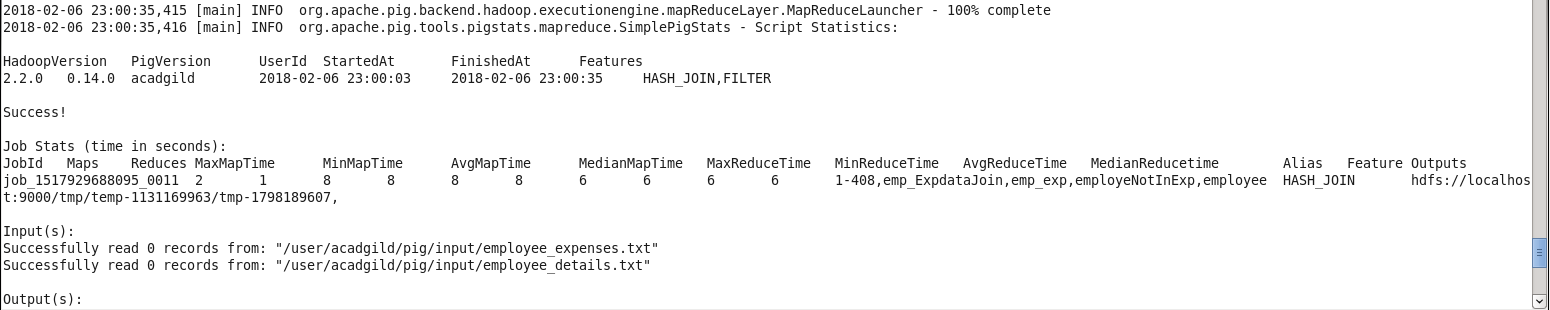
(d) List of employees (employee id and employee name) having entries in employee\_expenses

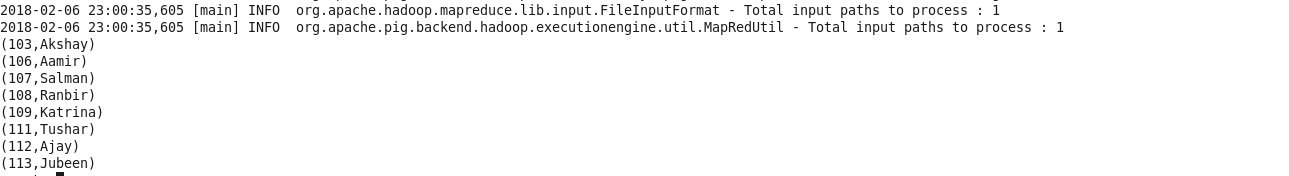
file.



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







--------------------------------------------------------------------------------------------------------------------

Use local mode while running Pig and write Pig Latin script

Assignment5.pig

---------------

employee = load '/user/acadgild/pig/input/employee\_details.txt' using PigStorage(',') as (emp\_id:int, emp\_name:chararray, emp\_salary:int, emp\_Rating:int);

emp\_exp = load '/user/acadgild/pig/input/employee\_expenses.txt' AS (emp\_id:int, emp\_expenses:int);

empLimit =LIMIT(ORDER employee BY emp\_Rating Desc, emp\_id Asc)5;

dump emplimit;

employee\_rating = FOREACH empLimit GENERATE emp\_id,emp\_name;

dump employee\_rating;

oddEmpId = FOREACH employee GENERATE emp\_id,(

CASE emp\_id % 2

WHEN 0 THEN -1

WHEN 1 THEN emp\_salary

END);

emp\_data\_filtered = FILTER oddEmpId BY (emp\_salary != -1);

emp\_join = JOIN emp\_data\_filtered by emp\_id, employee by emp\_id;

dump emp\_join

emp\_limit = FOREACH (LIMIT (ORDER emp\_join BY employee::emp\_salary Desc) 3)GENERATE employee::emp\_id, employee::emp\_name;

dump emp\_limit;

employeeJoinExp = JOIN emp\_exp BY emp\_id LEFT OUTER, employee by emp\_id;

dump employeeJoinExp

employeeElist = DISTINCT (FOREACH employeeJoinExp GENERATE emp\_exp::emp\_id, employee::emp\_name);

dump employeeElist;

employeeJoinMaxExp = JOIN employee BY emp\_id , emp\_exp by emp\_id;

emp\_MaxExp = FOREACH (LIMIT (ORDER employeeJoinMaxExp BY emp\_expenses Desc, emp\_name Asc) 1) GENERATE employee::emp\_id, employee::emp\_name;

dump emp\_MaxExp

emp\_ExpdataJoin = FILTER (JOIN emp\_exp by emp\_ID RIGHT OUTER, employee by emp\_id) by (emp\_exp::emp\_id IS NULL)dump emp\_MaxExp

employeNotInExp = FOREACH emp\_ExpdataJoin GENERATE employee::emp\_id, employee::emp\_name;

DUMP employeNotInExp



