**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)**

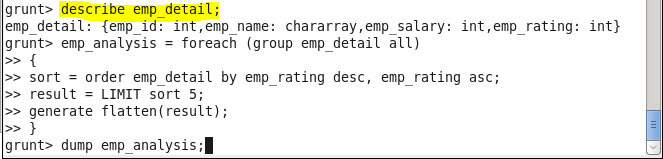
**Load the file from its location to the pig environment.**

****

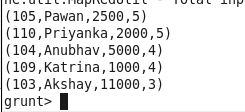
**Know the schema of emp\_detail using describe command.**

**Load the below command for the solution of given problem.**

1. **We group all together using group command.**
2. **Sort the name and rating on ascending and descending order respectively.**
3. **Limit the record to 5 so only 5 records get display.**
4. **Flatten the result so to get results in tuple and column wise.**

****

**Dump the emp\_analysis so to get the output.**

****

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

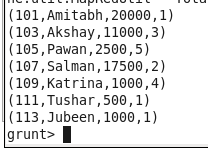
**is an odd number. (In case two employees have same salary, employee with name coming first**

**in dictionary should get preference)**

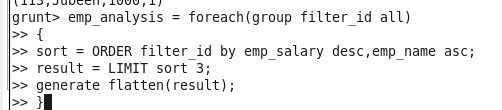
**Load the file from its location to the pig environment.**

****

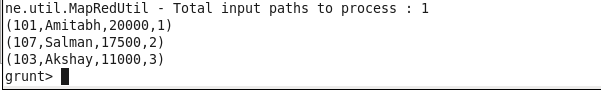
**Filter the records on basis of emp\_id which are odd**

****

**Load the below command for the solution of given problem.**

1. **We group all together using group command.**
2. **Sort the name and rating on ascending and descending order respectively.**
3. **Limit the record to 3 so only 3 records get display.**
4. **Flatten the result so to get results in tuple and column wise**

**Dump the emp\_analysis so to get the output**

****

**(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**

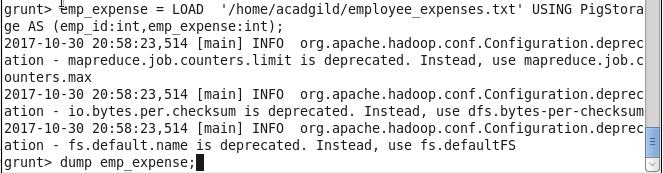
**preference)**

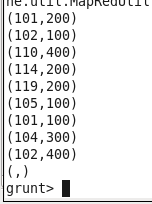
**Load the file from its location to the pig environment.**

****

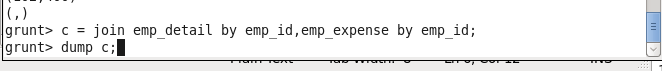
**Load the employee expense file also**

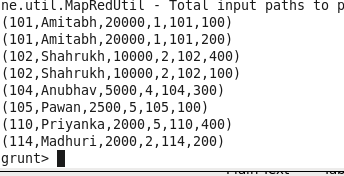
**Dump the file names ad emp\_expense.**

****

****

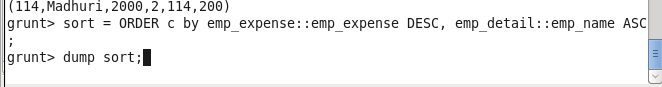
**Join the two table on the basis of there id**

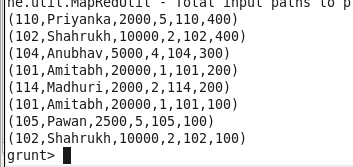
****

****

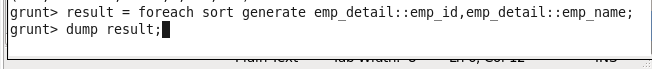
**Order the employee expense**

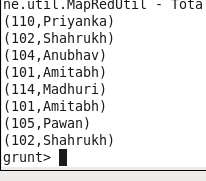
**And get the output.**

****

****

**Generate the emp name and emp id for the result**

****

****

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

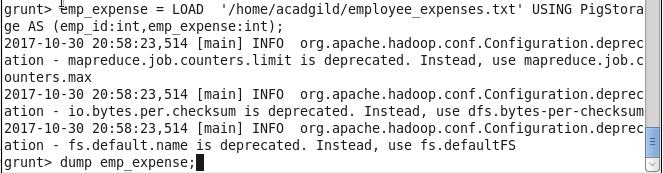
**file.**

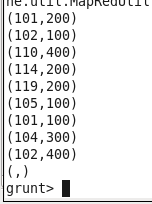
**Load the file from its location to the pig environment.**

****

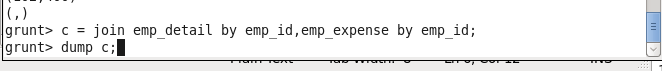
**Load the employee expense file also**

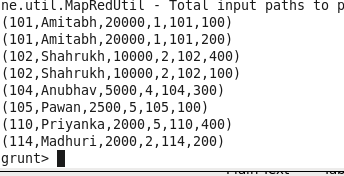
**Dump the file names ad emp\_expense.**

****

****

**Join the two table on the basis of there id**

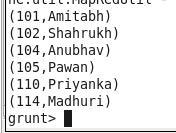
****

****

**Will get the value duplicate so use distinct to get the distinct value.**

****

****

****

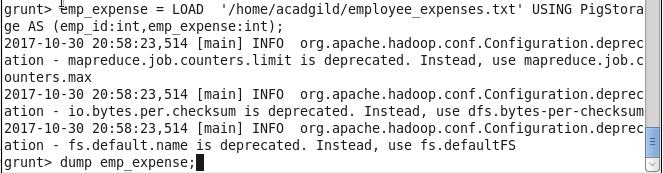
**(e) List of employees (employee id and employee name) having no entry in employee\_expenses**

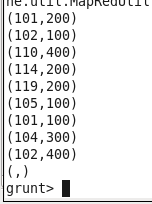
**Load the file from its location to the pig environment.**

****

**Load the employee expense file also**

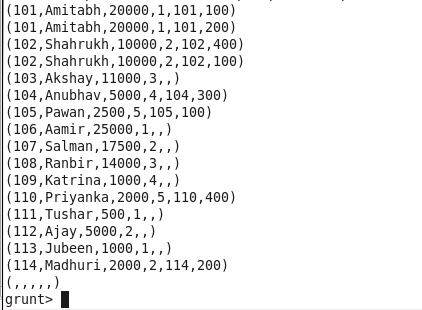
**Dump the file names ad emp\_expense.**

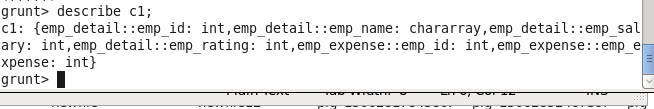
****

****

**Do the left outer join to get all the details of table emp\_detail and emp\_expense.**

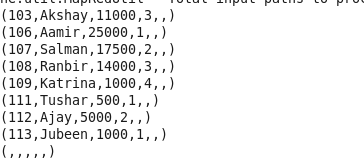
****

****

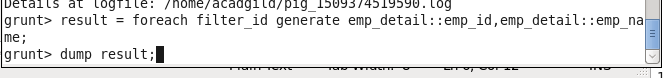
****

**Filter the record on basis of employee not having in expense table.and dump the table.**

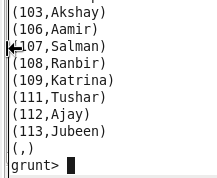
****

****

**Will generate a record with only name and id of employee with expense.**

****

**Will get the output**

****