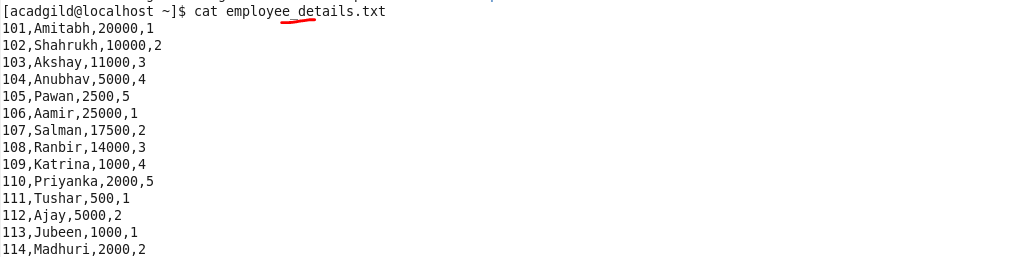
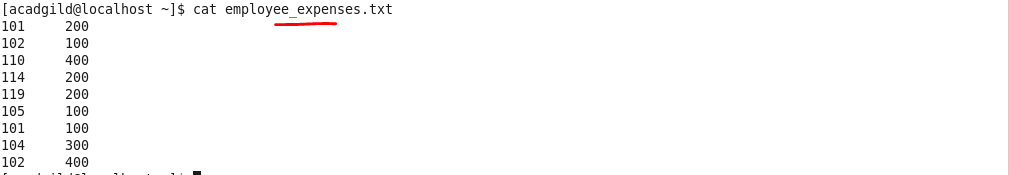
**ACD\_BDDOF\_Session\_2\_Assignment\_4\_Main**

**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:

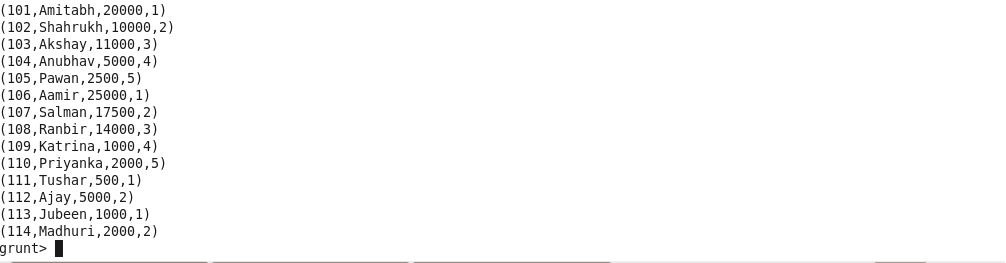




**Load the file to hdfs using PigStorage:**

Loaded the dataset and defined a schema for each fields in the file(which is separated by delimiter “,”) using PIG command



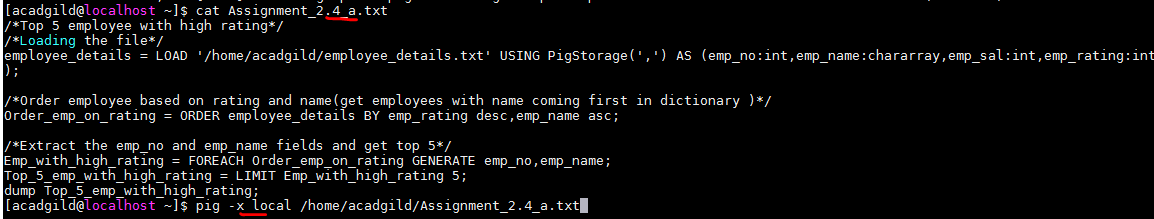


(a) Top 5 employees (employee id and employee name) with highest rating. (In case two

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

preference)

**Script:**

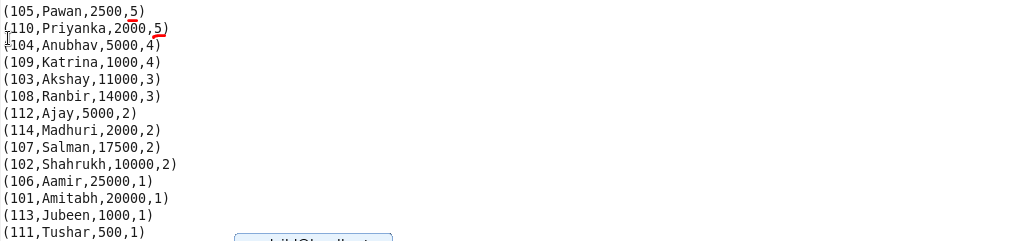


**Command i):**

Order employee details file based on **descending - emp\_rating**(highest rating) and **ascending – emp\_name** (employee with name coming first in dictionary should get

Preference)





**Final output:**

Top 5 employees (**employee id and employee name**) with highest rating.

**Command ii):**

Extracted only emp name and no for top 5 records.



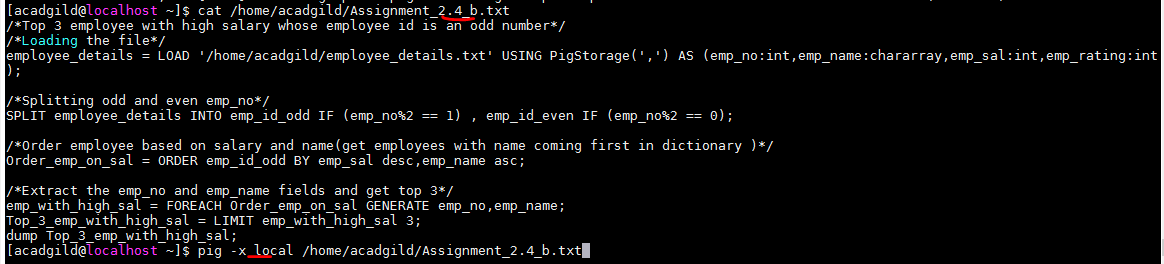


(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id

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

in dictionary should get preference)

**Script:**



**Command i):**

Splitting odd and even emp\_no



**Command ii):**

Order employee details file(where id is odd) based on **descending - emp\_salary**(highest salary) and **ascending – emp\_name** (employee with name coming first in dictionary should get

Preference)





**Final output:**

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

is an odd number.

**Command iii):**

Extracted only emp name and no for top 3 records.





(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)

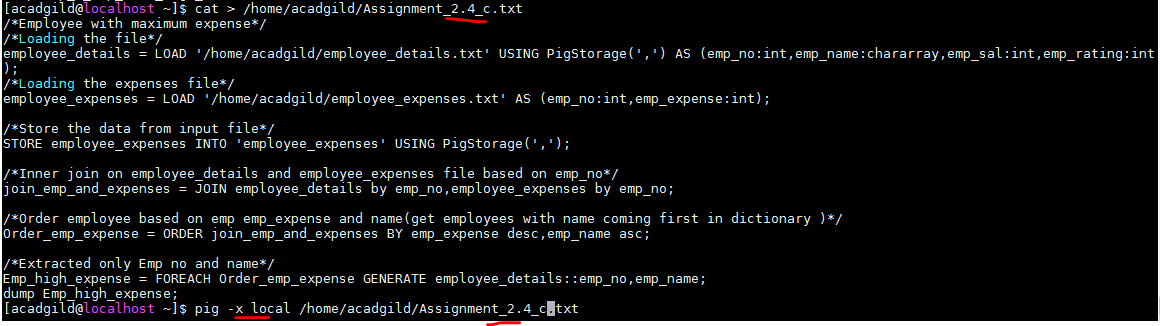
**Load the file to hdfs using PigStorage:**

Loaded the dataset and defined a schema for each fields in the file(which is separated by delimiter “,”) using PIG command





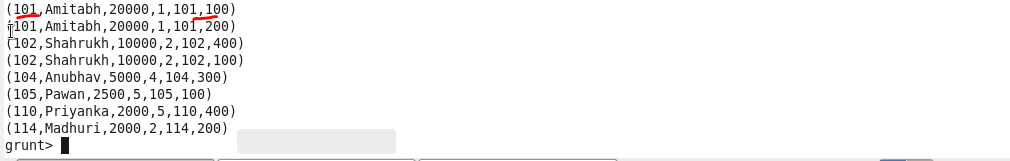
**Script:**



**Command i):**

Inner join on employee\_details and employee\_expenses file based on emp\_no

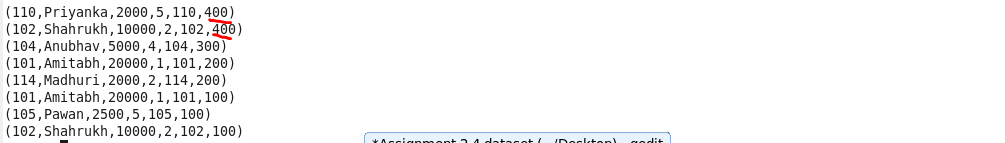




**Command ii):**

Order data based on **descending - emp\_expense** (highest salary) and **ascending – emp\_name** (employee with name coming first in dictionary should get Preference)





**Final output:**

Employee (employee id and employee name) with maximum expense

**Command iii):**

Extracted only Emp no and name

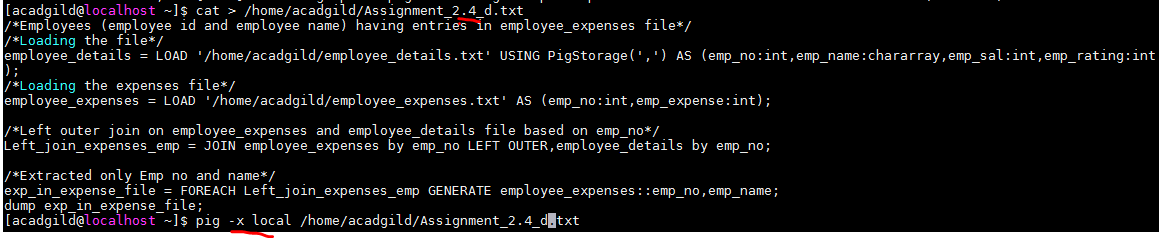




(d) List of employees (employee id and employee name) having entries in employee\_expenses

file.

**Script:**



**Command i):**

Left outer join on employee\_expenses and employee\_details file based on emp\_no



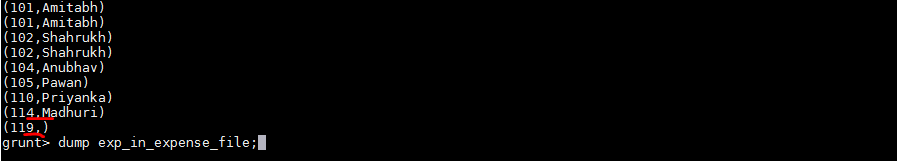


**Final output:**

**Command ii):**

Extracted only Emp no and name

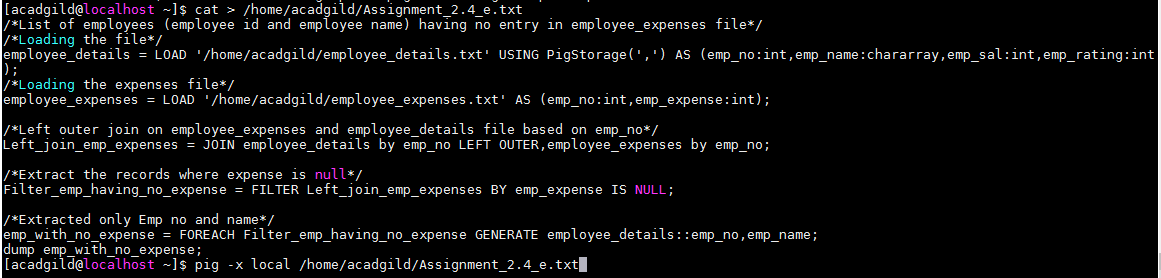




(e) List of employees (employee id and employee name) having no entry in employee\_expenses

file.

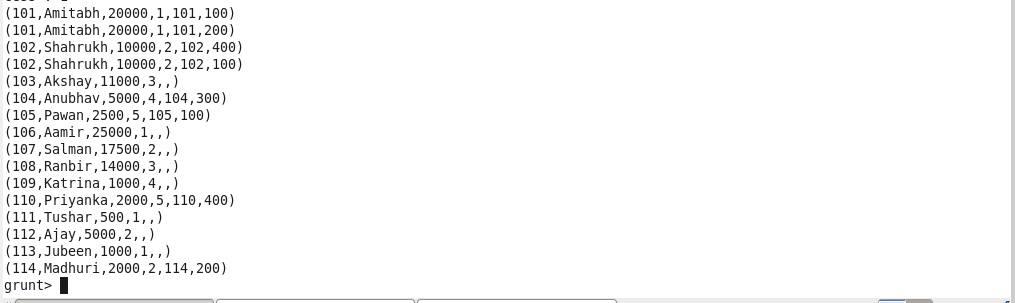
**Script:**



**Command i):**

Left outer join on employee\_expenses and employee\_details file based on emp\_no



**Command ii):**

Extract the records where expense is null**Final output:**

**Command iii):**

Extracted only Emp no and name.

