⇨ Task : 1

```
[acadgild@localhost pig]$ pig -x local
```

```
grunt> emp_details = LOAD 'pig/employee_details.txt' USING PigStorage (',') as (empid:Int,empname:chararray,salary:Int,rating
>> :Int);
2018-05-08 20:50:28,457 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-08 20:50:28,461 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt>
grunt>
```

emp_details = LOAD 'pig/employee_details.txt' USING PigStorage (',') as (empid:Int,empname:chararray,salary:Int,rating:Int);

emp_expenses= LOAD ' pig /employee_expenses.txt' USING PigStorage(' ') AS(empId:int,empexpenses:int);

emp_Highest_rating = ORDER emp_details by rating desc, empname asc;

dump emp_Highest_rating;

emp_top5_sal = LIMIT emp_Highest_rating 5;

dump emp_top5_sal;

```
grunt> emp_Highest_rating = ORDER emp_details  by rating desc, empname asc;
```

```
grunt> dump emp_Highest_rating;
```

```
2018-05-08 22:58:54,550
(105,Pawan,2500,5)
(110,Priyanka,2000,5)
(104,Anubhav,5000,4)
(109,Katrina,1000,4)
(103,Akshay,11000,3)
(108,Ranbir,14000,3)
(112,Ajay,5000,2)
(114,Madhuri,2000,2)
(107,Salman,17500,2)
(102,Shahrukh,10000,2)
(106,Aamir,25000,1)
(101,Amitabh,20000,1)
(113,Jubeen,1000,1)
(111,Tushar,500,1)
```

```
Kshay,11000,3)
emp_top5_sal = LIMIT emp_Highest_rating 5;
```

```
> dump emp_top5_sal;
```

```
2018-05-08 22:53:38,312 [M
(105,Pawan,2500,5)
(110,Priyanka,2000,5)
(104,Anubhav,5000,4)
(109,Katrina,1000,4)
(103,Akshay,11000,3)
grunt>
```

⇨ Task : 2

emp_details = LOAD 'pig/employee_details.txt' USING PigStorage (',') as (empid:Int,empname:chararray,salary:Int,rating:Int);

emp_Highest_Salary = ORDER emp_details by salary desc,empname asc;

empid_oddF = FILTER emp_Highest_Salary BY $0%2==1;

emp_top3_sal = LIMIT empid_oddF 3;

```
grunt> dump emp_details;
```

```
grunt> emp_Highest_Salary  = ORDER emp_details by salary desc;
```

```
grunt> dump emp_Highest_Salary;
```

```
grunt> empid_oddF = FILTER emp_Highest_Salary BY $0%2==1;
```

```
grunt> dump empid_oddF;
```

```
grunt> emp_top3_sal = LIMIT empid_oddF 3;
```

```
grunt> dump emp_top3_sal;
```

```
2018-05-08 20:05:59,727 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total i
(101,Amitabh,20000,1)
(107,Salman,17500,2)
(103,Akshay,11000,3)
grunt>
```

Task 3:

emp_details= LOAD 'pig/employee_details.txt' USING PigStorage(',')
AS(empId:int,empName:chararray,Salary:int,rating:int);

```
emp_expenses= LOAD 'pig/employee_expenses.txt' USING PigStorage('\t')
AS(empId:int,empexpenses:int);

join_empdetails_expenses= JOIN emp_details BY empId, emp_expenses BY empId;

joined_empdata = FOREACH join_empdetails_expenses GENERATE emp_details::empId,
emp_details::empName, emp_expenses::empexpenses;

joined_empdata_group = GROUP joined_empdata by (empId,empName);

joined_empdata_sum = FOREACH joined_empdata_group GENERATE group,
SUM(joined_empdata.emp_expenses::empexpenses) as sum;

joined_data_order = ORDER joined_empdata_sum by sum DESC;

joined_data_limit = LIMIT joined_data_order 1;

result = FOREACH joined_data_limit GENERATE FLATTEN(group);

DUMP result;
```

```
grunt> emp_details= LOAD 'pig/employee_details.txt' USING PigStorage(',') AS(empId:int,empName:chararray,Salary:int,rating:int);
emp_expenses= LOAD 'pig/employee_expenses.txt' USING PigStorage('\t') AS(empId:int,empexpenses:int);
join_empdetails_expenses= JOIN emp_details BY empId, emp_expenses BY empId;
joined_empdata = FOREACH join_empdetails_expenses GENERATE emp_details::empId, emp_details::empName, emp_expenses::empexpenses;
joined_empdata_group = GROUP joined_empdata by (empId,empName);
joined_empdata_sum = FOREACH joined_empdata_group GENERATE group, SUM(joined_empdata.emp_expenses::empexpenses) as sum;
joined_data_order = ORDER joined_empdata_sum by sum DESC;
joined_data_limit = LIMIT joined_data_order 1;
result = FOREACH joined_data_limit GENERATE FLATTEN(group);

2018-05-15 22:06:48,992 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-15 22:06:48,992 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-05-15 22:06:49,881 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-15 22:06:49,884 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> DUMP result;
```

```
(102,Shahrukh)
grunt>
```

Task 4:

```
emp_details = LOAD 'pig/employee_details.txt' USING PigStorage (',') as
(empid:Int,empname:chararray,salary:Int,rating:Int);

emp_expenses = LOAD 'pig/employee_expenses.txt' USING PigStorage as (empid:Int,expenses:Int);

join_empdetails_expenses = JOIN emp_details by empid,emp_expenses by empid;

empid_names_output = FOREACH join_empdetails_expenses GENERATE
emp_details::empid,emp_details::empname;

result = DISTINCT empid_names_output;

dump result;
```

```
grunt> emp_details = LOAD 'pig/employee_details.txt' USING PigStorage (',') as (empid:Int,empname:chararray,salary:Int,rating:Int);
2018-05-15 19:10:38,578 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-15 19:10:38,578 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> emp_expenses = LOAD 'pig/employee_expenses.txt' USING PigStorage as (empid:Int,expenses:Int);
2018-05-15 19:10:39,076 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-15 19:10:39,077 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> join_empdetails_expenses = JOIN emp_details by empid,emp_expenses by empid;
grunt> empid_names_output = FOREACH join_empdetails_expenses GENERATE emp_details::empid,emp_details::empname;
grunt> result = DISTINCT empid_names_output;
grunt> dump result;
```

```
2018-05-15 19:10:45,0
(101,Amitabh)
(102,Shahrukh)
(104,Anubhav)
(105,Pawan)
(110,Priyanka)
(114,Madhuri)
```

Task 5:

emp_details = LOAD 'pig_input_assignment5/employee_details.txt' USING PigStorage (',') as (empid:Int,empname:chararray,salary:Int,rating

:Int);

emp_expenses = LOAD 'pig_input_assignment5/employee_expenses.txt' using PigStorage as(empid:int,expenses:int);

join_emp_det_expenses = JOIN emp_details by empid LEFT OUTER,emp_expenses by empid;

result = FILTER join_emp_det_expenses by emp_expenses::empid is null;

final_result = FOREACH result GENERATE emp_details::empid,emp_details::empname;

dump final_result;

```
grunt> emp_details = LOAD 'pig_input_assignment5/employee_details.txt' USING PigStorage (',') as (empid:Int,empname:chararray,salary:Int,rating
>> :Int);
2018-05-15 19:11:33,805 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-15 19:11:33,805 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> emp_expenses = LOAD 'pig_input_assignment5/employee_expenses.txt' using PigStorage as(empid:int,expenses:int);
2018-05-15 19:11:34,365 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2018-05-15 19:11:34,365 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> join_emp_det_expenses = JOIN emp_details by empid LEFT OUTER,emp_expenses by empid;
grunt> result = FILTER join_emp_det_expenses by emp_expenses::empid is null;
grunt> final_result = FOREACH result GENERATE emp_details::empid,emp_details::empname;
grunt> dump final_result;
```

```
2017-12-12 16:14
(103,Akshay)
(106,Aamir)
(107,Salman)
(108,Ranbir)
(109,Katrina)
(111,Tushar)
(112,Ajay)
(113,Jubeen)
grunt>
```