**Assignment 5.1**

pig -x local

emp\_details = LOAD '/home/acadgild/hadoop/employee\_details.txt' USING PigStorage(',') AS (emp\_id:int, emp\_name:chararray, emp\_salary:int, emp\_rating:int);

DESCRIBE emp\_details

DUMP emp\_details

emp\_expenses = LOAD '/home/acadgild/hadoop/employee\_expenses.txt' AS (emp\_id:int, expenses:int);

DESCRIBE emp\_expenses

DUMP emp\_expenses

**Problem 5.1-a**

order\_by\_emp\_name\_rating = ORDER emp\_details BY emp\_rating ASC,emp\_name ASC;

DESCRIBE order\_by\_emp\_name\_rating

DUMP order\_by\_emp\_name\_rating

top\_5\_emp = limit order\_by\_emp\_name\_rating 5

DESCRIBE top\_5\_emp

DUMP top\_5\_emp

**Problem 5.1-b**

filtered\_emp\_id = FILTER emp\_details by emp\_id%2!=0;

DESCRIBE filtered\_emp\_id

DUMP filtered\_emp\_id

order\_by\_emp\_name\_salary = ORDER filtered\_emp\_id BY emp\_salary DESC,emp\_name ASC;

DESCRIBE order\_by\_emp\_name\_salary

DUMP order\_by\_emp\_name\_salary

top\_3\_emp = limit order\_by\_emp\_name\_salary 3

DUMP top\_3\_emp

**Problem 5.1-c**

joined\_data = join emp\_details by emp\_id , emp\_expenses by emp\_id;

DESCRIBE joined\_data

DUMP joined\_data

grunt> order\_by\_emp\_name\_expenses = ORDER joined\_data BY expenses DESC, emp\_name ASC;

DUMP order\_by\_emp\_name\_expenses

emp\_max\_expenses = limit order\_by\_emp\_name\_expenses 1;

DUMP emp\_max\_expenses

**Problem 5.1-d**

joined\_data = join emp\_details by emp\_id , emp\_expenses by emp\_id;

DESCRIBE joined\_data

DUMP joined\_data

**Problem 5.1-e**

full\_outer\_joined\_data = join emp\_details by emp\_id full outer, emp\_expenses by emp\_id;

DESCRIBE full\_outer\_joined\_data

DUMP full\_outer\_joined\_data

filtered\_emp\_id\_name = FILTER full\_outer\_joined\_data BY emp\_expenses::emp\_id is null;

final\_list = FOREACH flitered\_emp\_id\_name GENERATE $0,$1;

DUMP final\_list