

Database Project (SWE3033) (Fall 2024)

Homework #3 (50pts, Due date: 10/08)

Student ID: _____2018310773_____

Student Name: _____임승재_____

Instruction: We provided all codes and datasets. Implement two types of clauses, WHERE and JOIN. Please write your code to get correct result.

Submit two files as follows:

- 'DBP_HW3_STUDENTID.zip'
 - Code: DBP_HW3_STUDENTID_Code
 - Document: DBP_HW3_STUDENTID.pdf

1. [20pts] Given the dataset of employees and departments that you used in the lab session, write the code to find the names of all employees who are not in the Finance department and have a salary greater than or equal to 5000, and show your result.

Answer: Enter your code and result here. You must show your result, either an image or text.

```
mapper:
#!/usr/bin/env python
# coding: utf-8

import sys

for line in sys.stdin:
    line = line.strip()
    tuple_list = line.split(",")

    if len(tuple_list) == 11:
        ename = tuple_list[1] + ' ' + tuple_list[2]
        salary = tuple_list[-4]
        deptno = tuple_list[-1]
        value = ','.join(['emp', ename, salary])
        print('{0}\t{1}'.format(deptno, value))
    else:
        dname = tuple_list[1]
        deptno = tuple_list[0]
        value = ','.join(['dept', dname])
        print('{0}\t{1}'.format(deptno, value))

reducer:
#!/usr/bin/env python3
# * coding: utf-8 *
```

```

import sys
prev_deptno = None
dname = None
enames = []

for key_value in sys.stdin:
    key_value = key_value.strip()
    deptno, table_value = key_value.split("\t")
    table, value = table_value.split(",", maxsplit=1)

    if prev_deptno == deptno:
        if table == 'emp':
            enames.append(value)
        else:
            dname = value

    else:
        if prev_deptno and dname != 'Finance' and enames:
            for ename in enames:
                print("%s, %s"%(ename, dname))

            dname = None
            enames = []

        if table == 'emp':
            enames = [value]
        else:
            dname = value
        prev_deptno = deptno

if prev_deptno and dname != 'Finance' and enames:
    for ename in enames:
        print("%s, %s"%(ename, dname))

```

Result:

```

William Gietz,8300, Accounting
Shelley Higgins,12008, Accounting
Pat Fay,6000, Marketing
Michael Hartstein,13000, Marketing
Den Raphaely,11000, Purchasing
Susan Mavris,6500, Human Resources
Kevin Mourgous,5800, Shipping
Shanta Vollman,6500, Shipping
Payam Kaufling,7900, Shipping
Adam Fripp,8200, Shipping
Matthew Weiss,8000, Shipping
Bruce Ernst,6000, IT
Alexander Hunold,9000, IT
Hermann Baer,10000, Public Relations
Steven King,24000, Executive
Neena Kochhar,17000, Executive
Lex De Haan,17000, Executive

```

2. **[30pts]** We have provided the code with inner join implemented in the lab session. Modify this code to implement left join and right join and show your result.

Answer: Enter your code and result here. You must show your result, either an image or text.

(1) Left join

Mapper:

```
#!/usr/bin/env python
# coding: utf-8

import sys

for line in sys.stdin:
    line = line.strip()
    tuple_list = line.split(",")

    if len(tuple_list) == 11:
        ename = tuple_list[1] + ' ' + tuple_list[2]
        salary = tuple_list[-4]
        deptno = tuple_list[-1]
        value = ','.join(['emp', ename, salary])
        print('{0}¥t{1}'.format(deptno, value))
    else:
        dname = tuple_list[1]
        deptno = tuple_list[0]
        value = ','.join(['dept', dname])
        print('{0}¥t{1}'.format(deptno, value))
```

Reducer:

```
#!/usr/bin/env python3
# _* coding: utf-8 _*

import sys
prev_deptno = None
dname = None
enames = []

for key_value in sys.stdin:
    key_value = key_value.strip()
    deptno, table_value = key_value.split("¥t")
    table, value = table_value.split(",", maxsplit=1)

    if prev_deptno == deptno:
        if table == 'emp':
            enames.append(value)
        else:
            dname = value
    else:
        if prev_deptno and enames:
            for ename in enames:
                if dname:
                    print("%s, %s"%(ename, dname))
                else:
                    print("%s, %s"%(ename, "NULL"))

            dname = None
            enames = []
```

```

        if table == 'emp':
            enames = [value]
        else:
            dname = value
        prev_deptno = deptno

    if prev_deptno and enames:
        for ename in enames:
            if dname:
                print("%s, %s"%(ename, dname))
            else:
                print("%s, %s"%(ename, "NULL"))

```

Result:

Jennifer Whalen,4400, Administration
 Luis Popp,6900, Finance
 Jose Manuel Urman,7800, Finance
 Nancy Greenberg,12008, Finance
 Daniel Faviert,9000, Finance
 John Chen,8200, Finance
 Ismael Sciarra,7700, Finance
 Shelley Higgins,12008, Accounting
 William Gietz,8300, Accounting
 Michael Hartstein,13000, Marketing
 Pat Fay,6000, Marketing
 Den Raphaely,11000, Purchasing
 Alexander Khoo,3100, Purchasing
 Shelli Baida,2900, Purchasing
 Sigal Tobias,2800, Purchasing
 Guy Himuro,2600, Purchasing
 Karen Colmenares,2500, Purchasing
 Susan Mavris,6500, Human Resources
 Joshua Patel,2500, Shipping
 John Seo,2700, Shipping
 Stephen Stiles,3200, Shipping
 Renske Ladwig,3600, Shipping
 Hazel Philtanker,2200, Shipping
 Ki Gee,2400, Shipping
 Michael Rogers,2900, Shipping
 Jason Mallin,3300, Shipping
 TJ Olson,2100, Shipping
 James Marlow,2500, Shipping
 Mozhe Atkinson,2800, Shipping
 Laura Bissot,3300, Shipping
 Steven Markle,2200, Shipping
 James Landry,2400, Shipping
 Irene Mikkilineni,2700, Shipping
 Julia Nayer,3200, Shipping
 Kevin Mourgous,5800, Shipping
 Shanta Vollman,6500, Shipping
 Payam Kaufling,7900, Shipping
 Adam Fripp,8200, Shipping

| |
|--------------------------------------|
| Donald OConnell,2600, Shipping |
| Matthew Weiss,8000, Shipping |
| Douglas Grant,2600, Shipping |
| Alexander Hunold,9000, IT |
| Bruce Ernst,6000, IT |
| Valli Pataballa,4800, IT |
| Diana Lorentz,4200, IT |
| David Austin,4800, IT |
| Hermann Baer,10000, Public Relations |
| Neena Kochhar,17000, Executive |
| Lex De Haan,17000, Executive |
| Steven King,24000, Executive |

(2) Right join

Mapper:

```
#!/usr/bin/env python
# coding: utf-8

import sys

for line in sys.stdin:
    line = line.strip()
    tuple_list = line.split(",")

    if len(tuple_list) == 11:
        ename = tuple_list[1] + ' ' + tuple_list[2]
        salary = tuple_list[-4]
        deptno = tuple_list[-1]
        value = ','.join(['emp', ename, salary])
        print('{0}¥t{1}'.format(deptno, value))
    else:
        dname = tuple_list[1]
        deptno = tuple_list[0]
        value = ','.join(['dept', dname])
        print('{0}¥t{1}'.format(deptno, value))
```

Reducer:

```
#!/usr/bin/env python3
# _ * coding: utf-8 _ *

import sys
prev_deptno = None
dname = None
enames = []

for key_value in sys.stdin:
    key_value = key_value.strip()
    deptno, table_value = key_value.split("¥t")
    table, value = table_value.split(",", maxsplit=1)

    if prev_deptno == deptno:
        if table == 'emp':
            enames.append(value)
        else:
            dname = value
    else:
        if prev_deptno and dname:
            if enames:
                for ename in enames:
                    print("%s, %s"%(ename, dname))
            else:
                print("%s, %s"("NULL", dname))

        dname = None
        enames = []
```

```

        if table == 'emp':
            enames = [value]
        else:
            dname = value
            prev_deptno = deptno

if prev_deptno and dname:
    if enames:
        for ename in enames:
            print("%s, %s"%(ename, dname))
    else:
        print("%s, %s"%"(NULL", dname))

```

Result:

Jennifer Whalen,4400, Administration
 Luis Popp,6900, Finance
 Jose Manuel Urman,7800, Finance
 Nancy Greenberg,12008, Finance
 Daniel Faviat,9000, Finance
 John Chen,8200, Finance
 Ismael Sciarra,7700, Finance
 Shelley Higgins,12008, Accounting
 William Gietz,8300, Accounting
 NULL, Treasury
 NULL, Corporate Tax
 NULL, Control And Credit
 NULL, Shareholder Services
 NULL, Benefits
 NULL, Manufacturing
 NULL, Construction
 NULL, Contracting
 Michael Hartstein,13000, Marketing
 Pat Fay,6000, Marketing
 NULL, Operations
 NULL, IT Support
 NULL, NOC
 NULL, IT Helpdesk
 NULL, Government Sales
 NULL, Retail Sales
 NULL, Recruiting
 NULL, Payroll
 Den Raphaely,11000, Purchasing
 Alexander Khoo,3100, Purchasing
 Shelli Baida,2900, Purchasing
 Sigal Tobias,2800, Purchasing
 Guy Himuro,2600, Purchasing
 Karen Colmenares,2500, Purchasing
 Susan Mavris,6500, Human Resources
 Joshua Patel,2500, Shipping
 John Seo,2700, Shipping
 Stephen Stiles,3200, Shipping
 Renske Ladwig,3600, Shipping
 Hazel Philtanker,2200, Shipping

Ki Gee,2400, Shipping
Michael Rogers,2900, Shipping
Jason Mallin,3300, Shipping
TJ Olson,2100, Shipping
James Marlow,2500, Shipping
Mozhe Atkinson,2800, Shipping
Laura Bissot,3300, Shipping
Steven Markle,2200, Shipping
James Landry,2400, Shipping
Irene Mikkilineni,2700, Shipping
Julia Nayer,3200, Shipping
Kevin Mourgos,5800, Shipping
Shanta Vollman,6500, Shipping
Payam Kaufling,7900, Shipping
Adam Fripp,8200, Shipping
Donald OConnell,2600, Shipping
Matthew Weiss,8000, Shipping
Douglas Grant,2600, Shipping
Alexander Hunold,9000, IT
Bruce Ernst,6000, IT
Valli Pataballa,4800, IT
Diana Lorentz,4200, IT
David Austin,4800, IT
Hermann Baer,10000, Public Relations
NULL, Sales
Neena Kochhar,17000, Executive
Lex De Haan,17000, Executive
Steven King,24000, Executive