

Ομάδα Εργασίας :
ΕΜΜΑΝΟΥΗΛ-ΧΡΗΣΤΟΣ ΛΙΛΛΗΣ
Π19091
ΣΤΕΦΑΝΟΣ ΔΟΣΚΑΡΗΣ
Π19046
ΑΛΕΞΑΝΔΡΟΣ ΜΠΑΤΣΙΚΩΣΤΑΣ
Π19113

Γενική εξήγηση :

Στο πρόγραμμα έχει προστεθεί η λειτουργία των query τύπου :

```
select (aggregate function),col,... from table where (condition) group  
by col... having (condition)
```

το where και το having είναι προαιρετικά.

Ξεκινώντας από το mdb.py όταν δοθεί ένα τέτοιο query στον τερματικό
το πρόγραμμα συνεχίζει όπως και πριν μέχρι να καλεστεί η

select από το database.py, με την διαφορά ότι το dic έχει επιπλέον το
group by και το having για keys και τα αντίστοιχα values.

Η select έχει 3 επιπλέον ορίσματα : το group_by που αν είναι true
γεμίζει το group_by_col_list και το aggr_func_list και καλείται η

select_where από το table.py. Στην select_where δίνονται επιπλέον τα
group_by_col_list, aggr_func_list και τα υπόλοιπα καινούργια ορίσματα

της select, το having_cond. Με τις κατάλληλες αλλαγές στην
select_where και στην parse_condition και σε συνεργασία με τις
καινούργιες

συναρτήσεις(`group_by_func`,`count`,`sum`,`avg`,`min`,`max`) τροποποιείται το `table.data` και ύστερα δημιουργείται το `dict` που χρησιμοποιείται στην `show()` για να εμφανιστούν τα αποτελέσματα του query.

Αλλαγές σε κάθε αρχείο :

Αλλαγές στο αρχείο `mdb.py` :

Στην συνάρτηση `create_query_plan` έχει προστεθεί συνθήκη για να χωρίζονται τα περιέχόμενα του `group by` από το `having`.

Στην `interpret_query` έχουν προστεθεί τα `group_by` και `having` ως keywords.

Στο τέλος στο `result.show()` έχει προστεθεί έλεγχος για το αν χρησιμοποιείται το `group_by`, για να αναλάβει την περίπτωση που αναφέρεται στα σχόλια.

Αλλαγές στο αρχείο `database.py` :

Η συνάρτηση `select` έχει δύο επιπλέον ορίσματα (`group by`,`having_cond`) και εξηγείται σε σχόλια η λειτουργία τους(στο `else` του `self.has_index(table_name)`) και αν το `group_by` αληθεύει οδηγεί στην συνάρτηση `select_where` του `table.py`.

Αλλαγές στο αρχείο `table.py` :

Η συνάρτηση `select_where` έχει τροποποιηθεί αρκετά για να συμπεριλάβει την περίπτωση του `group_by` με ή χωρίς το `having`. Αρχικά έχει

3 επιπλέον ορίσματα

(`group_by_col_list`,`aggr_func_list`,`having_cond`). Τροποποιείται η λίστα `return_columns` ώστε αν υπάρχει `group_by`

να έχει στην αρχή της την aggregate function και ύστερα ότι υπάρχει στο group_by_col_list.

`πχ select count(month),year,days group by year,days`

```
aggr_func_list = [count(month)]
```

```
group_by_col_list = [year,days]
```


```
return_columns = [count(month),year,days]
```

Λόγω της αλλαγής του `return_columns` έχει τροποποιηθεί και η `parse_condition` για να αναγνωρίζει συνθήκες τύπου `count(month) > 1`.

Τέλος οι συναρτήσεις `group_by_func`, `count`, `min`, `max`, `sum`, `avg` έχουν υλοποιηθεί και εξηγείται σε σχόλια η λειτουργία τους.

Σκρινσοτ εκτέλεσης :

```
(base) ubuntu@ubuntu2004:~$ cd miniDB
(base) ubuntu@ubuntu2004:~/miniDB$ conda activate mdb
(mdb) ubuntu@ubuntu2004:~/miniDB$ DB=smbd python3.9 mdb.py
```



The ASCII art logo features the word "minidB" in a stylized font where each letter is composed of vertical bars of varying heights. To the right of the letters is a small icon resembling a database cylinder or a storage unit. Below the main text, it says "2021 - v3.2".

```
(smbd)> select a,b,c from tests2
bb
{'select': None, 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
['select', 'a,b,c', 'from', 'tests2', ';']
['select', 'from', ';']
[0, 2, 4]
{'select': 'a,b,c', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'a,b,c', 'from': 'tests2', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
['table name', 'pid', 'mode']
```

```
ubuntu@ubuntu2004: ~/miniDB
['select', 'a,b,c', 'from', 'tests2', ';']
['select', 'from', ';']
[0, 2, 4]
{'select': 'a,b,c', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'a,b,c', 'from': 'tests2', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
['table_name', 'pid', 'mode']
[<class 'str'>, <class 'int'>, <class 'str'>]
{'_name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None}
['a', 'b', 'c']
[<class 'int'>, <class 'int'>, <class 'int'>]
{'_name': 'tests2', 'column_names': ['a', 'b', 'c'], 'columns': [[], [], []], 'a': [], 'b': [], 'c': [], 'column_types': [<class 'int'>, <class 'int'>, <class 'int'>], 'data': [[5, 3, 4], [5, 8, 4], [3, 8, 4]], 'pk_idx': None, 'pk': None}


| a (int) | b (int) | c (int) |
|---------|---------|---------|
| 5       | 3       | 4       |
| 5       | 8       | 4       |
| 3       | 8       | 4       |


(smdb)>
```

```
ubuntu@ubuntu2004: ~/miniDB
return getattr(db, action)(*dic.values())
File "/home/ubuntu/miniDB/miniDB/database.py", line 381, in select
    table = self.tables[table_name]._select_where(columns, condition, order_by, desc, top_k, group_by_col_list, aggr_func_list, having_cond)
File "/home/ubuntu/miniDB/miniDB/table.py", line 238, in _select_where
    return_cols = [self.column_names.index(col.strip()) for col in return_column
s.split(',')]
File "/home/ubuntu/miniDB/miniDB/table.py", line 238, in <listcomp>
    return_cols = [self.column_names.index(col.strip()) for col in return_column
s.split(',')]
ValueError: 'count(a)' is not in list

(smdb)> select count(a),b,c from tests2 where b > 3 group by b,c
bb
{'select': None, 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
['select', 'count', '(', 'a', ')', ', b,c', 'from', 'tests2', 'where', 'b', '>', '3', 'group', 'by', 'b,c', ';']
['select', 'from', 'where', 'group by', ';']
[0, 6, 8, 13, 15]
{'select': 'count ( a ) ,b,c', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
```

```
ubuntu@ubuntu2004: ~/miniDB
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': 'b,c', 'having': None}
['table_name', 'pid', 'mode']
[<class 'str'>, <class 'int'>, <class 'str'>]
{'_name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None}
['a', 'b', 'c']
[<class 'int'>, <class 'int'>, <class 'int'>]
['count ( a ) ', 'b', 'c']
[[5, 3, 4], [5, 8, 4], [3, 8, 4]]
[[8, 4]]
{'_name': 'tests2', 'column_names': ['count ( a ) ', 'b', 'c'], 'columns': [[], [], []], 'a': [], 'b': [], 'c': [], 'column_types': [<class 'int'>, <class 'int'>, <class 'int'>], 'data': [[2, 8, 4]], 'pk_idx': None, 'pk': None, 'data_before_print': [[5, 3, 4], [5, 8, 4], [3, 8, 4]], 'group_by_value_lists': [], 'comb_list': [[2, 8, 4]], 'group_by_distinct': [8, 4], 'count_of_col_in_combs': [2]}
count ( a ) (int)      b (int)      c (int)
-----
                2          8          4

(smdb)>
```

```
ubuntu@ubuntu2004: ~/miniDB
by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': 'b,c', 'having': None}
['table_name', 'pid', 'mode']
[<class 'str'>, <class 'int'>, <class 'str'>]
{'_name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None}
['a', 'b', 'c']
[<class 'int'>, <class 'int'>, <class 'int'>]
['count ( a ) ', 'b', 'c']
[[5, 3, 4], [5, 8, 4], [3, 8, 4]]
[[8, 4]]
{'_name': 'tests2', 'column_names': ['count ( a ) ', 'b', 'c'], 'columns': [[], [], []], 'a': [], 'b': [], 'c': [], 'column_types': [<class 'int'>, <class 'int'>, <class 'int'>], 'data': [[2, 8, 4]], 'pk_idx': None, 'pk': None, 'data_before_print': [[5, 3, 4], [5, 8, 4], [3, 8, 4]], 'group_by_value_lists': [], 'comb_list': [[2, 8, 4]], 'group_by_distinct': [8, 4], 'count_of_col_in_combs': [2]}
count ( a ) (int)      b (int)      c (int)
-----
                2          8          4

(smdb)> select count(a),b,c from tests2 where b > 3 group by b,c having count(a)
= 2
```

```
ubuntu@ubuntu2004: ~/miniDB
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': 'b,c', 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': 'b,c', 'having': 'count ( a ) = 2'}
['table_name', 'pid', 'mode']
[<class 'str'>, <class 'int'>, <class 'str'>]
{'_name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None}
['a', 'b', 'c']
[<class 'int'>, <class 'int'>, <class 'int'>]
['count ( a ) ', 'b', 'c']
[[5, 3, 4], [5, 8, 4], [3, 8, 4]]
[[8, 4]]
{'_name': 'tests2', 'column_names': ['count ( a ) ', 'b', 'c'], 'columns': [[], [], []], 'a': [], 'b': [], 'c': [], 'column_types': [<class 'int'>, <class 'int'>, <class 'int'>], 'data': [[2, 8, 4]], 'pk_idx': None, 'pk': None, 'data_before_print': [[5, 3, 4], [5, 8, 4], [3, 8, 4]], 'group_by_value_lists': [], 'comb_list': [[2, 8, 4]], 'group by distinct': [8, 4], 'count of col in combs': [2]}
count ( a ) (int)      b (int)      c (int)
-----
                2          8          4

(smdb)> 
```

```
ubuntu@ubuntu2004: ~/miniDB
= 2
bb
{'select': None, 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count', 'from': '(', 'a', ')', 'b,c', 'from': 'tests2', 'where': 'b', '>', '3', 'group', 'by', 'b,c', 'having': 'count', '(', 'a', ')', 'a', '= 2', ')'}
{'select': 'from', 'where': 'group by', 'having': ')'}
[0, 0, 0, 10, 15, 22]
{'select': 'count ( a ) ,b,c', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': 'b,c', 'having': None}
{'select': 'count ( a ) ,b,c', 'from': 'tests2', 'where': 'b > 3 group', 'order by': None, 'top': None, 'group by': 'b,c', 'having': 'count ( a ) = 2'}
{'table_name', 'pid', 'mode'}
[<class 'str'>, <class 'int'>, <class 'str'>]
{'_name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None}
['a', 'b', 'c']
[<class 'str'>, <class 'int'>, <class 'int'>]
['count ( a ) ', 'b', 'c']
[[5, 3, 4], [5, 8, 4], [3, 8, 4]]
[[8, 4]]
{'_name': 'tests2', 'column_names': ['count ( a ) ', 'b', 'c'], 'columns': [[], [], []], 'a': [], 'b': [], 'c': [], 'column_types': [<class 'int'>, <class 'int'>, <class 'int'>], 'data': [[2, 8, 4]], 'pk_idx': None, 'pk': None, 'data_before_print': [[5, 3, 4], [5, 8, 4], [3, 8, 4]], 'group_by_value_lists': [], 'comb_list': [[2, 8, 4]], 'group by distinct': [8, 4], 'count of col in combs': [2]}
count ( a ) (int)      b (int)      c (int)
-----
                2          8          4

(smdb)> select an,bn from tests3
bb
{'select': None, 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'an,bn', 'from': 'tests3', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'from', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
[0, 2, 4]
{'select': 'an,bn', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'select': 'an,bn', 'from': 'tests3', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None}
{'table_name', 'pid', 'mode'}
[<class 'str'>, <class 'int'>, <class 'str'>]
{'_name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None}
['an', 'bn']
[<class 'str'>, <class 'int'>]
{'_name': 'tests3', 'column_names': ['an', 'bn'], 'columns': [[], [], []], 'an': [], 'bn': [], 'column_types': [<class 'str'>, <class 'int'>], 'data': [['love', 23], ['apple', 25], ['mainou', 27], ['p-phthikos', 27]], 'pk_idx': None, 'pk': None}
an (str)      bn (int)
-----
love          23
apple         25
mainou        27
p-phthikos    27

(smdb)> 
```



```
ubuntu@ubuntu2004:~/miniDB
[0, 2, 4]
['select': 'year,month', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,month', 'from': 'test', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['table_name', 'pid', 'mode']
['class 'str', 'class 'int', 'class 'str']
['name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None]
['year', 'month']
['class 'int', 'class 'int']
['name': 'test', 'column_names': ['year', 'month'], 'columns': [[], [], 'year': [], 'month': []], 'column_types': [<class 'int', <class 'int'>], 'data': [[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]], 'pk_idx': None, 'pk': None, 'group_by_value_lists': [], 'comb_list': [[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]], 'group_by_distinct': [2003]]
year (int) month (int)
-----
11      2001
5        2003
2003     5
2001     10
2001     20

(sndb)> select year,avg(month) from test group by year
bb
['select': None, 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,avg', '(', 'month', ')', 'from', 'test', 'group', 'by', 'year', ';']
['select', 'from', 'group by', ':']
[0, 5, 8, 10]
['select': 'year,avg ( month )', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,avg ( month )', 'from': 'test group', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,avg ( month )', 'from': 'test group', 'where': None, 'order by': None, 'top': None, 'group by': 'year', 'having': None]
['table_name', 'pid', 'mode']
['class 'str', 'class 'int', 'class 'str']
['name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None]
['year', 'month']
['class 'int', 'class 'int']
['avg ( month )', 'year']
[[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]]
[[11], [5], [2003], [2001]]
[2001, 2003, 5, 15]
[[2001, 11], [2003, 5], [5, 2003], [15, 2001]]
['name': 'test', 'column_names': ['avg ( month )', 'year'], 'columns': [[], [], 'year': [], 'month': []], 'column_types': [<class 'int', <class 'int'>], 'data': [[2001, 11], [2003, 5], [5, 2003], [15, 2001]], 'pk_idx': None, 'pk': None, 'group_by_value_lists': [], 'comb_list': [[2001, 11], [2003, 5], [5, 2003], [15, 2001]], 'group_by_distinct': [2001], 'data_before_print': [[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]]
avg ( month ) (int) year (int)
-----
2001     11
2003     5
5        2003
15       2001

(sndb)>
```

```
ubuntu@ubuntu2004:~/miniDB
[0, 2, 4]
['select': 'year,month', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,month', 'from': 'test', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['table_name', 'pid', 'mode']
['class 'str', 'class 'int', 'class 'str']
['name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None]
['year', 'month']
['class 'int', 'class 'int']
['name': 'test', 'column_names': ['year', 'month'], 'columns': [[], [], 'year': [], 'month': []], 'column_types': [<class 'int', <class 'int'>], 'data': [[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]], 'pk_idx': None, 'pk': None, 'group_by_value_lists': [], 'comb_list': [[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]], 'group_by_distinct': [2003]]
year (int) month (int)
-----
11      2001
5        2003
2003     5
2001     10
2001     20

(sndb)> select year,avg(month) from test group by year
bb
['select': None, 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,avg', '(', 'month', ')', 'from', 'test', 'group', 'by', 'year', ';']
['select', 'from', 'group by', ':']
[0, 5, 8, 10]
['select': 'year,avg ( month )', 'from': None, 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,avg ( month )', 'from': 'test group', 'where': None, 'order by': None, 'top': None, 'group by': None, 'having': None]
['select': 'year,avg ( month )', 'from': 'test group', 'where': None, 'order by': None, 'top': None, 'group by': 'year', 'having': None]
['table_name', 'pid', 'mode']
['class 'str', 'class 'int', 'class 'str']
['name': 'meta_locks', 'column_names': ['pid'], 'columns': [[], [], []], 'table_name': [], 'pid': [], 'mode': [], 'column_types': [<class 'int'>], 'data': [], 'pk_idx': None, 'pk': None]
['year', 'month']
['class 'int', 'class 'int']
['avg ( month )', 'year']
[[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]]
[[11], [5], [2003], [2001]]
[2001, 2003, 5, 15]
[[2001, 11], [2003, 5], [5, 2003], [15, 2001]]
['name': 'test', 'column_names': ['avg ( month )', 'year'], 'columns': [[], [], 'year': [], 'month': []], 'column_types': [<class 'int', <class 'int'>], 'data': [[2001, 11], [2003, 5], [5, 2003], [15, 2001]], 'pk_idx': None, 'pk': None, 'group_by_value_lists': [], 'comb_list': [[2001, 11], [2003, 5], [5, 2003], [15, 2001]], 'group_by_distinct': [2001], 'data_before_print': [[11, 2001], [5, 2003], [2003, 5], [2001, 10], [2001, 20]]
avg ( month ) (int) year (int)
-----
2001     11
2003     5
5        2003
15       2001

(sndb)>
```