

ЛАБОРАТОРНАЯ РАБОТА №11

Тема: Раздел GROUP BY предложения SELECT. Исследование особенностей составления SQL запросов с использованием группировки.

Цель: Пробрести навыки составления SQL запросов использующих группировку записей.

В данной лабораторной работе используются база данных HR, созданная в рамках первой лабораторной работы. Также используется таблица TB_ELEKTROSTAL_2018, которая создана и наполнена данными в процессе выполнения лаб. работы №.1.

Дополнительно используется удаленная база данных проекта stackoverflow.com. Структура этой базы данных описана в методических указаниях к выполнению лабораторной работы № 3. Напомним, адрес ресурса <http://data.stackexchange.com/stackoverflow/query/new>.

В SQL предусматривается использование следующих итоговых (агрегатных) функций:

- COUNT
- SUM
- MAX
- MIN
- AVG

Функция COUNT используется для подсчета строк или значений в столбцах отличных от значения NULL. При использовании с опцией DISTINCT функция COUNT подсчитает количество различных значений.

Функция SUM используется для подсчета суммы значений в столбе для заданной группы строк.

Функция AVG используется для подсчета среднего значения в столбе для заданной группы строк.

Функция MAX используется для поиска максимального значения в столбце для заданной группы строк.

Функция MIN используется для поиска минимального значения в столбе для заданной группы строк.

Группирование данных осуществляется при помощи выражения GROUP BY в операторе SELECT. Ключевое слово GROUP BY должно следовать за выражением WHERE и предшествовать ключевому слову ORDER BY. Правильный формат оператора SQL следующий:

```
SELECT колонка1, колонка2, ..., колонка N  
FROM таблица1, таблица2...  
GROUP BY колонка1, колонка2  
HAVING условия  
ORDER BY колонка1, колонка2
```

В выражении ключевого слова GROUP BY можно использовать только выбранные колонки (колонки из списка полей, или колонки, используемые при построении фильтра в разделе WHERE предложения SELECT. Если имя колонки отсутствует в списке полей или выражении фильтра, то это имя нельзя использовать для задания условия группировки после ключевого слова GROUP BY. С другой стороны, имена колонок списка полей или имена колонок, участвующих в формировании фильтра, **должны** присутствовать в списке задания условия группировки после ключевого слова GROUP BY.

Для решения задач, когда условие отбора налагается на группу, а не на каждую строку предназначено ключевое слово HAVING. Это ключевое слово играет ту же роль для групп, что и ключевое слово WHERE для каждой строки. Другими словами, WHERE задает условия для отбора значений из выбранных столбцов, HAVING задает условия для групп, создаваемых с помощью GROUP BY.

Примеры выполнения запросов с использованием итоговых функций

Подсчитать количество строк в таблице POSTS БД stackoverflow.

```
Select count (*) from posts
```

Подсчитать количество пользователей, создававших посты в таблице posts с использованием опции DISTINCT.

```
select count(distinct OwnerUserId) from posts
```

Подсчитать количество абонентов, оплативших коммунальные услуги.

```
Select count(distinct shifrabo) from kvit
```

Подсчитать среднюю зарплату работников (БД HR).

```
Select avg(salary) from employees
```

Найти размер максимальной зарплаты работников.

```
Select max(salary) from employees
```

Найти размер минимальной зарплаты работников.

```
Select min(salary) from employees
```

Примеры выполнения запросов, осуществляющих группирование данных

Вывести список улиц, сгруппированный по типам (таблица tb_elekrostal_2018).

```
select  namestr  
        , typstr  
        from tb_elektrostal_2018  
        group by typstr,namestr
```

Вывести список, содержащий для каждого типа улиц их количество.

```
select  typstr  
        , count(distinct namestr)  from tb_elektrostal_2018  
        group by typstr order by 2
```

В условиях предыдущего задания вывести только те типы улиц, которым соответствует не менее 15 наименований улиц. Для решения задачи, можно использовать ключевое слово HAVING.

```
select  typstr  
        , count(distinct namestr)  from tb_elektrostal_2018  
        group by typstr  
        having count(distinct name)>14  
        order by 2
```

Упорядочить результаты предыдущего запроса по убыванию количества наименований улиц. Для решения этой проблемы используем предложение ORDER BY. Данное предложение может вместо имен колонок использовать их номера, которые определяются порядком расположения колонок после ключевого слова select.

```

select  typstr
        , count(distinct namestr) from tb_elektrostal_2018
        group by typstr
        having count(*)>14
        order by 2 desc

```

Операция группировки часто используется для поиска экстремальных (максимальных или минимальных) значений. Для решения задач на поиск экстремума используется сочетание сортировки и отбора первой записи. Пусть требуется найти тип улицы, для которого в таблице имеется наибольшее количество названий улиц

```

select typstr,cnt from (
select rownum rn, typstr, cnt from (
select typstr
        , count(distinct namestr) cnt from
tb_elektrostal_2018
        group by typstr
        having count(distinct namestr)>14
        order by 2 desc) o) o1
where rn=1

```

Порядок выполнения работы

Порядок выполнения работы.

Составить следующие запросы

- Составить запрос к БД HR. Результат запроса должен содержать одно число – количество департаментов по данным только таблицы employees.
- Составить запрос к БД HR. Результат запроса должен содержать две колонки. Первая колонка название должности. Вторая колонка – средняя зарплата у работников, занимающих эту должность.
- Составить запрос к БД HR. Результат запроса должен содержать две колонки. Первая колонка название должности. Вторая колонка – суммарный фонд оплаты труда по этой должности. Этот фонд для каждой

должности определяется суммированием окладов всех работников, занимающих эту должность.

- Составить запрос к БД HR. Результат запроса должен содержать три колонки. Первая колонка название должности. Вторая колонка – суммарный фонд оплаты труда по этой должности. Этот фонд для каждой должности определяется суммированием окладов всех работников, занимающих эту должность. Третья колонка – расчетный фонд оплаты труда работников этой должности. Расчетный фонд определяется умножением расчетного среднего оклада должности на количество работников этой должности. Расчетный средний оклад определяется как среднее арифметическое максимального и минимального должностного оклада. Максимальный и минимальный должностной оклад определить по таблице JOBS.
- В условиях предыдущего запроса вывести должность, у которой наибольшее отклонение расчетного фонда оплаты труда от расчетного фонда оплаты труда.
- Составить запрос к БД stackoverflow. Результат запроса должен содержать две колонки. Первая колонка – имя пользователя. Вторая колонка - количество постов этого пользователя.
- Составить запрос к БД stackoverflow. Результат запроса должен содержать имя пользователя создавшего больше всего постов в 2017 году.
- Составить запрос с использованием таблицы tb_elektrostal_2018. Вывести адрес дома, в котором наибольшее количество жителей.
- Составить запрос с использованием таблицы tb_elektrostal_2018. Вывести название и тип улицы, на которой расположено наибольшее количество домов.

Содержание отчета:

1. Тема, цель лабораторной работы.
2. Примеры выполнения запросов к базе данных.

3. Составленные согласно заданию запросы и скриншоты полученных результатов.

5. Выводы.

Контрольные вопросы:

1. Какие виды итоговых функций Вы знаете?
2. Расскажите, как осуществляется группирование данных.
3. Объясните, в чем отличие в использовании ключевых слов GROUP BY и HAVING.
4. Составьте правильный формат оператора SQL для группирования данных.
5. Может ли использоваться группировка без итоговых функций?
6. Дайте характеристику запросам поиска экстремальных значений с использованием группирования данных.
7. Объясните особенности использования вложенных запросов в последнем примере данных методических указаний. (Запрос на поиск типа улиц с максимальным количеством наименований).
8. Дайте описание особенностей использования единичных полей и итоговых (агрегатных функций).
9. Опишите назначение раздела HAVING предложения SELECT.

Файл: Лабораторная работа 11
Каталог: E:\Projects\documents\md\predmety\BD_2018
Шаблон: C:\Users\sss\AppData\Roaming\Microsoft\Шаблоны\Normal.dotm
Заголовок:
Содержание:
Автор: sss
Ключевые слова:
Заметки:
Дата создания: 17.11.2018 17:02:00
Число сохранений: 100
Дата сохранения: 17.11.2018 21:55:00
Сохранил: sss
Полное время правки: 287 мин.
Дата печати: 17.11.2018 21:56:00
При последней печати
 страниц: 6
 слов: 1 212 (прибл.)
 знаков: 6 911 (прибл.)