Chapter 7: GROUP BY

Results of a SELECT query can be grouped by one or more columns using the `GROUP BY` statement: all results with the same value in the grouped columns are aggregated together. This generates a table of partial results, instead of one result. GROUP BY can be used in conjunction with aggregation functions using the `HAVING` statement to define how non-grouped columns are aggregated.

Section 7.1: Basic GROUP BY example

It might be easier if you think of GROUP BY as "for each" for the sake of explanation. The query below:

```
SELECT EmpID, SUM (MonthlySalary)
FROM Employee
GROUP BY EmpID
```

is saying:

"Give me the sum of MonthlySalary's **for each** EmpID"

So if your table looked like this:

```
+-----+-------------+
|EmpID|MonthlySalary|
+-----+-------------+
|1    |200          |
+-----+-------------+
|2    |300          |
+-----+-------------+
```

Result:

```
+-+---+
|1|200|
+-+---+
|2|300|
+-+---+
```

Sum wouldn't appear to do anything because the sum of one number is that number. On the other hand if it looked like this:

```
+-----+-------------+
|EmpID|MonthlySalary|
+-----+-------------+
|1    |200          |
+-----+-------------+
|1    |300          |
+-----+-------------+
|2    |300          |
+-----+-------------+
```

Result:

```
```

```
+-+---+
|1|500|
+-+---+
|2|300|
+-+---+
```

Then it would because there are two EmpID 1's to sum together.

Section 7.2: Filter GROUP BY results using a HAVING clause

A HAVING clause filters the results of a GROUP BY expression. Note: The following examples are using the Library example database.

Examples:

Return all authors that wrote more than one book (live example).

```
SELECT
  a.Id,
  a.Name,
  COUNT(*) BooksWritten
FROM BooksAuthors ba
  INNER JOIN Authors a ON a.id = ba.authorid
GROUP BY
  a.Id,
  a.Name
HAVING COUNT(*) > 1    -- equals to HAVING BooksWritten > 1
;
```

Return all books that have more than three authors (live example).

```
SELECT
  b.Id,
  b.Title,
  COUNT(*) NumberOfAuthors
FROM BooksAuthors ba
  INNER JOIN Books b ON b.id = ba.bookid
GROUP BY
  b.Id,
  b.Title
HAVING COUNT(*) > 3    -- equals to HAVING NumberOfAuthors > 3
;
```

Section 7.3: USE GROUP BY to COUNT the number of rows for each unique entry in a given column

Let's say you want to generate counts or subtotals for a given value in a column.

Given this table, "Westerosians":

Name	GreatHouseAllegience
Arya	Stark
Cercei	Lannister
Myrcella	Lannister
Yara	Greyjoy
Catelyn	Stark

Sansa	Stark

Without GROUP BY, COUNT will simply return a total number of rows:

```
SELECT Count(*) Number_of_Westerosians
FROM Westerosians
```

returns...

Number_of_Westerosians
6

But by adding GROUP BY, we can COUNT the users for each value in a given column, to return the number of people in a given Great House, say:

```
SELECT GreatHouseAllegience House, Count(*) Number_of_Westerosians
FROM Westerosians
GROUP BY GreatHouseAllegience
```

returns...

House	Number_of_Westerosians
Stark	3
Greyjoy	1
Lannister	2

It's common to combine GROUP BY with ORDER BY to sort results by largest or smallest category:

```
SELECT GreatHouseAllegience House, Count(*) Number_of_Westerosians
FROM Westerosians
GROUP BY GreatHouseAllegience
ORDER BY Number_of_Westerosians Desc
```

returns...

House	Number_of_Westerosians
Stark	3
Lannister	2
Greyjoy	1

Section 7.4: ROLAP aggregation (Data Mining)

Description

The SQL standard provides two additional aggregate operators. These use the polymorphic value "ALL" to denote the set of all values that an attribute can take. The two operators are:

- `with data cube` that it provides all possible combinations than the argument attributes of the clause.
- `with roll up` that it provides the aggregates obtained by considering the attributes in order from left to right compared how they are listed in the argument of the clause.

SQL standard versions that support these features: 1999,2003,2006,2008,2011.

Examples

Consider this table:

Food	Brand	Total_amount
Pasta	Brand1	100
Pasta	Brand2	250
Pizza	Brand2	300

With cube

```
select Food,Brand,Total_amount
from Table
group by Food,Brand,Total_amount with cube
```

Food	Brand	Total_amount
Pasta	Brand1	100
Pasta	Brand2	250
Pasta	ALL	350
Pizza	Brand2	300
Pizza	ALL	300
ALL	Brand1	100
ALL	Brand2	550
ALL	ALL	650

With roll up

```
select Food,Brand,Total_amount
from Table
group by Food,Brand,Total_amount with roll up
```

Food	Brand	Total_amount
Pasta	Brand1	100
Pasta	Brand2	250
Pizza	Brand2	300
Pasta	ALL	350
Pizza	ALL	300
ALL	ALL	650