

# SQL 基礎(5) GROUP BY 句 - グループ化と集計

作成日: 2016/02/18

作成者: 西村

## 更新履歴

| 更新日        | 更新概要   | 作業者 |
|------------|--------|-----|
| 2016/02/18 | ・ 新規作成 | 西村  |
|            | ・      |     |
|            | ・      |     |
|            | ・      |     |
|            | ・      |     |

## はじめに

この資料では、GROUP BY 句によるグループ化と、データの合計などを取る方法について簡単に説明します。

- ・ GROUP BY 句 (グループ化)
- ・ 集約関数 (合計など)
- ・ HAVING 句 (グループ化されたデータの抽出)

## サンプルのデータ

---

この資料では、下記のテーブルをもとに各クエリの結果がどうなるかを示します。

[ 社員テーブル ]

| id | 姓  | 名  | 性別 | 部署 id |
|----|----|----|----|-------|
| 1  | 山田 | 太郎 | 男  | 1     |
| 2  | 福岡 | 幸子 | 女  | 2     |
| 3  | 東京 | 次郎 | 男  | 3     |
| 4  | 青森 | 花子 | 女  | 1     |
| 5  | 岩手 | 誠司 | 男  | 1     |

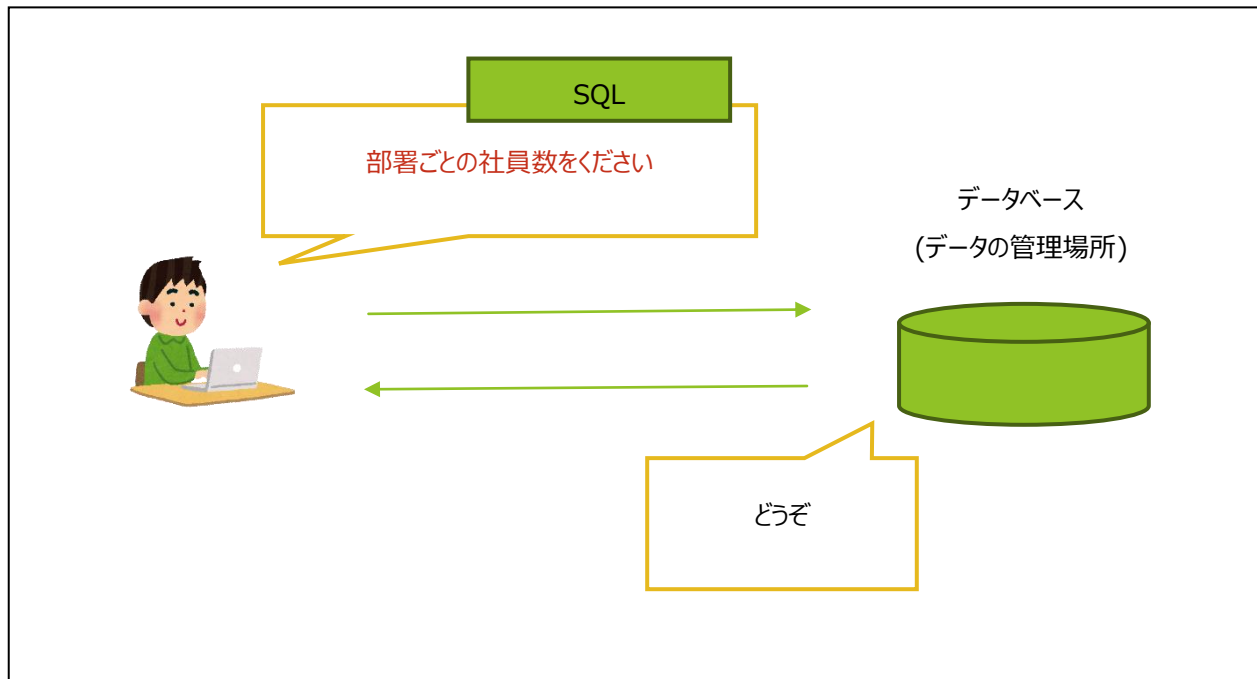
[ 給与テーブル ]

| id | 社員 id | 給与年  | 給与月 | 給与      |
|----|-------|------|-----|---------|
| 1  | 1     | 2016 | 4   | 150000  |
| 2  | 2     | 2016 | 4   | 160000  |
| 3  | 3     | 2016 | 4   | 170000  |
| 4  | 4     | 2016 | 4   | 180000  |
| 5  | 5     | 2016 | 4   | 190000  |
| 6  | 1     | 2016 | 5   | 140000  |
| 7  | 2     | 2016 | 5   | 170000  |
| 8  | 3     | 2016 | 5   | 160000  |
| 9  | 4     | 2016 | 5   | 190000  |
| 10 | 5     | 2016 | 5   | 1000000 |

## GROUP BY 句 ?

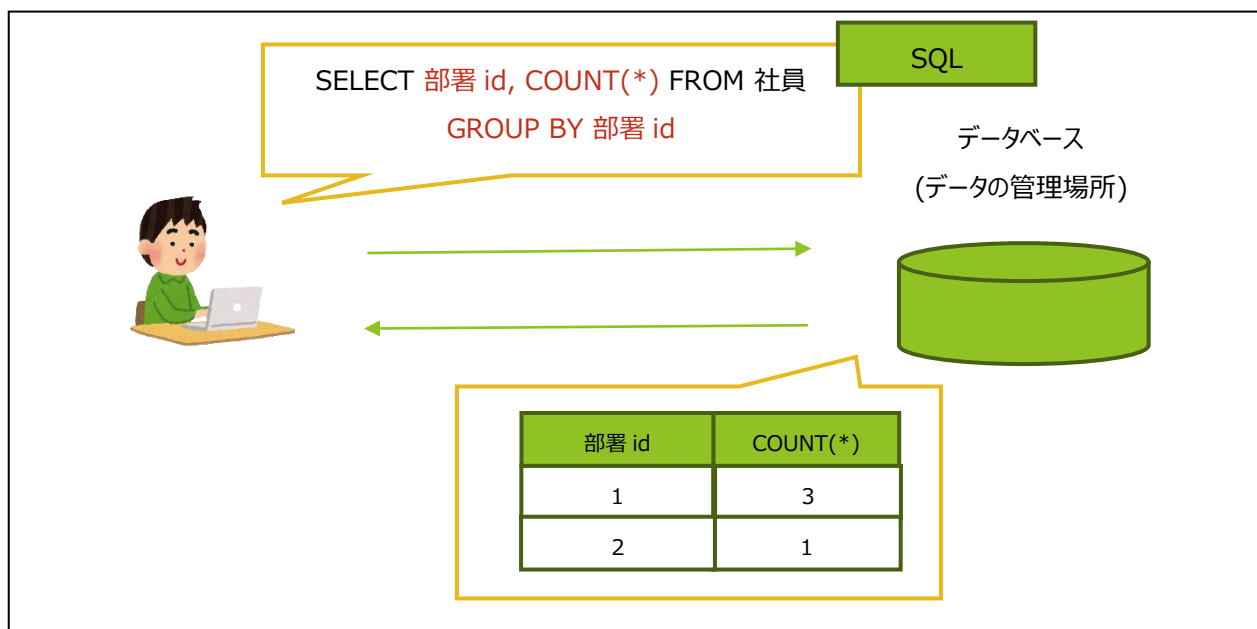
GROUP BY 句(GROUP BY clause)は、データベース内のデータをグループ化する機能です。

たとえば「部署ごとの人数」「月ごとの売上」など、グループごと(○○ごと)の集計をしたいときに使います。



(データを取得する際のイメージ)

実際にはこうなります。



基本的な書き方は下記ようになります。

```
SELECT 列名や集約関数
FROM テーブル名
GROUP BY グループ条件
```

「グループ条件」は、「まとめる条件」です。基本的には列を指定することが多いです。

列を指定すると、その指定した列の値が同じデータ同士が1つのグループになります。

※集約関数はあとで説明します。

※前に説明した WHERE や ORDER BY も使えます。GROUP BY の位置は WHERE の後、ORDER BY の前です。位置を間違えるとエラーになります。

部署 id でグループ化した場合、下記のように id [ 1, 4, 5 ] [ 2 ] [ 3 ] がグループ化されるイメージです。

| Id | 姓  | 名  | 性別 | 部署 id |
|----|----|----|----|-------|
| 1  | 山田 | 太郎 | 男  | 1     |
| 2  | 福岡 | 幸子 | 女  | 2     |
| 3  | 東京 | 次郎 | 男  | 3     |
| 4  | 青森 | 花子 | 女  | 1     |
| 5  | 岩手 | 誠司 | 男  | 1     |

このグループ化した中から合計や件数を取っていく形になります。

「社員テーブルから部署 id ごとに社員数(件数)を取得したい」という場合だと下記ようになります。

```
SELECT 部署 id, COUNT(*)  
FROM 社員  
GROUP BY 部署 id
```

日本語にするとこんなイメージです。

データを選択→ 部署 id, 部署ごとの件数  
このテーブルから→ 社員  
この値によってグループ化→ 部署 id

結果は下記ようになります。

| Id | 姓  | 名  | 性別 | 部署 id |
|----|----|----|----|-------|
| 1  | 山田 | 太郎 | 男  | 1     |
| 2  | 福岡 | 幸子 | 女  | 2     |
| 3  | 東京 | 次郎 | 男  | 3     |
| 4  | 青森 | 花子 | 女  | 1     |
| 5  | 岩手 | 誠司 | 男  | 1     |

↓

| 部署 id | COUNT(*) |
|-------|----------|
| 1     | 3        |
| 2     | 1        |
| 3     | 1        |

<http://sqlfiddle.com/#!9/ba981/1>

## 集約関数？

集約関数は、データをまとめるための機能です。Excel を使ったことがある人は、Excel の SUM 関数(指定したセルの値の合計を取る)などを思い浮かべてもらえるとわかりやすいと思います。

集約関数には下記のようなものがあります。

| 集約関数       | 意味         |
|------------|------------|
| COUNT( 式 ) | データの件数を数える |
| SUM( 式 )   | データの合計を取る  |
| AVG( 式 )   | データの平均を取る  |
| MIN( 式 )   | 最小のデータを取る  |
| MAX( 式 )   | 最大のデータを取る  |

※ 式には列名や計算式が入ります。

※ COUNT はだいたい「COUNT(\*)」が多いです。列名を指定すると、「NULL を除いたカウント」になるためです。

(例) 給与テーブルから年月ごとの最大の給与を取る

```
SELECT 給与年, 給与月, MAX(給与)
FROM 給与
GROUP BY 給与年, 給与月
```

※ GROUP BY も ORDER BY などと同じように複数の条件を指定できます。

[ 給与テーブル ]

| Id | 社員 id | 給与年  | 給与月 | 給与      |
|----|-------|------|-----|---------|
| 1  | 1     | 2016 | 4   | 150000  |
| 2  | 2     | 2016 | 4   | 160000  |
| 3  | 3     | 2016 | 4   | 170000  |
| 4  | 4     | 2016 | 4   | 180000  |
| 5  | 5     | 2016 | 4   | 190000  |
| 6  | 1     | 2016 | 5   | 140000  |
| 7  | 2     | 2016 | 5   | 170000  |
| 8  | 3     | 2016 | 5   | 160000  |
| 9  | 4     | 2016 | 5   | 190000  |
| 10 | 5     | 2016 | 5   | 1000000 |

↓

| 給与年  | 給与月 | MAX( 給与 ) |
|------|-----|-----------|
| 2016 | 4   | 190000    |
| 2016 | 5   | 1000000   |

MIN にすると最小の給与(15 万, 14 万), SUM にすると合計(85 万, 166 万)、AVG にすると平均(17 万, 33.2 万)になります。

<http://sqlfiddle.com/#!9/ba981/2>

## GROUP BYした SELECT で選択できるもの

GROUP BYした SELECT で選択できるものは基本的には「GROUP BYした列か集約関数」だけです。

たとえば、「部署 id でグループ化して部署 id を選択」は○ですが、「部署 id でグループ化して姓を選択」は×です。

```
SELECT 部署 id, 姓, COUNT(*)  
FROM 社員  
GROUP BY 部署 id
```

→ 姓でグループ化していないので何を返したらいいかデータベース側が特定できない

※ MySQL の場合はエラーにはなりません(MySQL は基本的にルーズでゆるふわです)が、何が取得できるかわからないので意図しない結果になることがあります。できるだけこのような書き方はやめたほうがよいです。

## GROUP BY で重複を除く

GROUP BY を応用してデータの重複を除いて結果を取得することもできます。

```
SELECT 部署 id  
FROM 社員  
GROUP BY 部署 id
```

→ サンプルデータだと 1, 2, 3 の 3 行が取得できる(本来は 1, 2, 3, 1, 1 ですが、グループ化されているので重複を除いたのと同じ結果になる)

<http://sqlfiddle.com/#!9/ba981/3>

ただ、単純に重複を除きたいだけなら「DISTINCT」という SELECT のオプション機能を使ったほうがわかりやすいです。

```
SELECT DISTINCT 部署 id FROM 社員
```

→ これも 1, 2, 3 の 3 行。「DISTINCT」は「別個の」という意味です。

<http://sqlfiddle.com/#!9/ba981/4>



## グループ化したデータをさらに絞り込む (HAVING 句)

「社員・年ごとに給与を合計して、合計給与が 30 万以上のデータだけ取得したい」など、グループ化したデータの行をさらに絞り込むには「HAVING」を使います。

基本的な書き方は下記ようになります。

```
SELECT 列名や集約関数  
FROM テーブル名  
GROUP BY グループ条件  
HAVING グループ抽出条件
```

HAVING は必ず GROUP BY のあとに書きます。

「社員・年ごとに給与を合計して、合計給与が 30 万以上のデータだけ取得」を SQL にすると下記ようになります。

```
SELECT 社員 id, 給与年, SUM(給与)  
FROM 給与  
GROUP BY 社員 id, 給与年  
HAVING SUM(給与) >= 300000
```

[ 給与テーブル ]

| Id | 社員 id | 給与年  | 給与月 | 給与      |
|----|-------|------|-----|---------|
| 1  | 1     | 2016 | 4   | 150000  |
| 2  | 2     | 2016 | 4   | 160000  |
| 3  | 3     | 2016 | 4   | 170000  |
| 4  | 4     | 2016 | 4   | 180000  |
| 5  | 5     | 2016 | 4   | 190000  |
| 6  | 1     | 2016 | 5   | 140000  |
| 7  | 2     | 2016 | 5   | 170000  |
| 8  | 3     | 2016 | 5   | 160000  |
| 9  | 4     | 2016 | 5   | 190000  |
| 10 | 5     | 2016 | 5   | 1000000 |

↓

(社員 id と給与年で GROUP BY して給与合計)

| 社員 id | 給与年  | SUM( 給与 ) |
|-------|------|-----------|
| 1     | 2016 | 290000    |
| 2     | 2016 | 330000    |
| 3     | 2016 | 330000    |
| 4     | 2016 | 370000    |
| 5     | 2016 | 1190000   |

↓

(「給与合計が 30 万以上」で HAVING)

| 社員 id | 給与年  | SUM( 給与 ) |
|-------|------|-----------|
| 2     | 2016 | 330000    |
| 3     | 2016 | 330000    |
| 4     | 2016 | 370000    |
| 5     | 2016 | 1190000   |

社員 id=1 の行は 30 万未満のため結果から取り除かれます。

<http://sqlfiddle.com/#!9/ba981/5>

## HAVINGとWHERE

HAVING では SUM() など以外に普通の列も抽出条件にすることができますが、普通の列での抽出条件は HAVING ではなく WHERE を使いましょう。

「グループ化してから絞り込む」より「絞り込んでからグループ化する」のほうがデータベースの負担が少なく速いので、WHERE ができることは WHERE で済ませたほうがよいです。

(×。グループ化してから絞り込んでいる)

```
SELECT 社員 id, 給与年, SUM(給与)
FROM 給与
GROUP BY 社員 id, 給与年
HAVING 給与年 >= 2016
```

(○。絞り込んでからグループ化している)

```
SELECT 社員 id, 給与年, SUM(給与)
FROM 給与
WHERE 給与年 >= 2016
GROUP BY 社員 id, 給与年
```

## 問題

---

1. 社員テーブルで性別ごとの社員数を取得するには何を使いますか？
2. 給与テーブルの全体の給与を取得するには何を使いますか？
3. 給与テーブルの年の社員ごとの最大給与を取得するには何を使いますか？
4. 社員テーブルで部署ごとの社員数を取得(社員数が3未満の部署は除く)するには何を使いますか？

(選択肢)

- A. SELECT 社員 id, 給与年, MAX(給与) FROM 給与  
GROUP BY 社員 id, 給与年
- B. SELECT SUM(給与) FROM 給与
- C. SELECT 性別, COUNT(\*) FROM 社員
- D. SELECT 性別, COUNT(\*) FROM 社員  
GROUP BY 性別
- E. SELECT 部署 id, COUNT(\*) FROM 社員  
GROUP BY 部署 id  
HAVING COUNT(\*) >= 3

上記がわかればだいたい理解ができていると思います。