

GROUP BY

 Cover

▼ CÔNG DỤNG

- Được sử dụng để **nhóm** các bản ghi có cùng giá trị và thực hiện các phép toán tổng hợp (như `COUNT`, `SUM`, `AVG`, `MIN`, `MAX`) cho mỗi nhóm.
- Nó giúp tóm tắt dữ liệu và dễ dàng phân tích các xu hướng hoặc mẫu chung.

▼ CÚ PHÁP

```
SELECT column1, column2, aggregate_function(column3)
FROM table_name
WHERE condition
GROUP BY column1, column2;
```

- `column1, column2` : Các cột mà bạn muốn nhóm.
- `aggregate_function(column3)` : Hàm tổng hợp áp dụng cho các nhóm, ví dụ: `COUNT`, `SUM`, `AVG`, `MAX`, `MIN`.
- `table_name` : Tên bảng chứa dữ liệu.
- `condition` : Điều kiện lọc (tùy chọn).

▼ VÍ DỤ

1. **Đếm số lượng nhân viên trong mỗi** `department_id`.

```
SELECT department_id, COUNT(*)
FROM employees
GROUP BY department_id;
```

2. **Tính lương trung bình (`avg_salary`) cho mỗi** `department_id`.

```
SELECT department_id, AVG(salary) AS avg_salary
FROM employees
GROUP BY department_id;
```

3. **Tính tổng lương (`total_salary`) và số lượng nhân viên (`num_employees`) trong mỗi** `department_id`.

```
SELECT department_id, SUM(salary) AS total_salary, COUNT(*) AS num_employees
FROM employees
GROUP BY department_id;
```

4. Đếm số lượng nhân viên cho mỗi cặp `department_id` và `job_id`.

```
SELECT department_id, job_id, COUNT(*)
FROM employees
GROUP BY department_id, job_id;
```

5. Tính lương trung bình (`avg_salary`) cho mỗi `department_id` của nhân viên có `job_id` là 'IT_PROG'.

```
SELECT department_id, AVG(salary) AS avg_salary
FROM employees
WHERE job_id = 'IT_PROG'
GROUP BY department_id;
```

6. Sử dụng với HAVING để lọc kết quả sau khi nhóm:

Ví dụ này sẽ chỉ trả về các `department_id` có lương trung bình (`avg_salary`) lớn hơn 50,000.

```
SELECT department_id, AVG(salary) AS avg_salary
FROM employees
GROUP BY department_id
HAVING AVG(salary) > 50000;
```

Bảng dữ liệu mẫu: `employees`

| employee_id | first_name | last_name | department_id | salary |
|-------------|------------|-----------|---------------|--------|
| 1 | John | Doe | 10 | 60000 |
| 2 | Jane | Smith | 20 | 75000 |
| 3 | Larry | Page | 10 | 65000 |
| 4 | Sergey | Brin | 30 | 80000 |
| 5 | Alan | Turing | 20 | 70000 |
| 6 | Ada | Lovelace | 30 | 90000 |
| 7 | Charles | Babbage | 10 | 72000 |
| 8 | Grace | Hopper | 20 | 80000 |
| 9 | Margaret | Hamilton | 30 | 85000 |
| 10 | Elon | Musk | 40 | 95000 |

▼ Tính tổng lương (`total_salary`) và số lượng nhân viên (`num_employees`) trong mỗi `department_id` từ bảng `employees`

```
SELECT department_id, SUM(salary) AS total_salary, COUNT(*) AS num_employees
FROM employees
GROUP BY department_id;
```

Kết quả

Kết quả của câu lệnh sẽ như sau:

| department_id | total_salary | num_employees |
|---------------|--------------|---------------|
| 10 | 197000 | 3 |
| 20 | 225000 | 3 |
| 30 | 255000 | 3 |
| 40 | 95000 | 1 |

Giải thích kết quả

- **department_id 10:** Có 3 nhân viên với tổng lương là 197000.
 - John Doe: 60000
 - Larry Page: 65000
 - Charles Babbage: 72000
- **department_id 20:** Có 3 nhân viên với tổng lương là 225000.
 - Jane Smith: 75000
 - Alan Turing: 70000
 - Grace Hopper: 80000
- **department_id 30:** Có 3 nhân viên với tổng lương là 255000.
 - Sergey Brin: 80000
 - Ada Lovelace: 90000
 - Margaret Hamilton: 85000
- **department_id 40:** Có 1 nhân viên với tổng lương là 95000.
 - Elon Musk: 95000

Tổng kết

Câu lệnh `SELECT department_id, SUM(salary) AS total_salary, COUNT(*) AS num_employees FROM employees GROUP BY department_id;` giúp bạn tính toán tổng lương và số lượng nhân viên trong mỗi phòng ban. Câu lệnh này sử dụng `GROUP BY department_id` để nhóm các hàng theo `department_id` và áp dụng các hàm tổng hợp `SUM` và `COUNT` để tính toán các giá trị tổng lương và số lượng nhân viên trong từng nhóm.

| employee_id | first_name | last_name | department_id | job_id | salary |
|-------------|------------|-----------|---------------|--------|--------|
| 1 | John | Doe | 10 | 101 | 60000 |
| 2 | Jane | Smith | 20 | 102 | 75000 |
| 3 | Larry | Page | 10 | 101 | 65000 |

| | | | | | |
|----|----------|----------|----|-----|-------|
| 4 | Sergey | Brin | 30 | 103 | 80000 |
| 5 | Alan | Turing | 20 | 104 | 70000 |
| 6 | Ada | Lovelace | 30 | 103 | 90000 |
| 7 | Charles | Babbage | 10 | 105 | 72000 |
| 8 | Grace | Hopper | 20 | 104 | 80000 |
| 9 | Margaret | Hamilton | 30 | 103 | 85000 |
| 10 | Elon | Musk | 40 | 106 | 95000 |

▼ **Đếm số lượng nhân viên cho mỗi cặp `department_id` và `job_id` từ bảng `employees`**

```
SELECT department_id, job_id, COUNT(*)
FROM employees
GROUP BY department_id, job_id;
```

Kết quả

Kết quả của câu lệnh sẽ như sau:

| department_id | job_id | count |
|---------------|--------|-------|
| 10 | 101 | 2 |
| 10 | 105 | 1 |
| 20 | 102 | 1 |
| 20 | 104 | 2 |
| 30 | 103 | 3 |
| 40 | 106 | 1 |

Giải thích kết quả

- **department_id 10, job_id 101:** Có 2 nhân viên.
 - John Doe: job_id 101
 - Larry Page: job_id 101
- **department_id 10, job_id 105:** Có 1 nhân viên.
 - Charles Babbage: job_id 105
- **department_id 20, job_id 102:** Có 1 nhân viên.
 - Jane Smith: job_id 102
- **department_id 20, job_id 104:** Có 2 nhân viên.
 - Alan Turing: job_id 104
 - Grace Hopper: job_id 104

- **department_id 30, job_id 103:** Có 3 nhân viên.
 - Sergey Brin: job_id 103
 - Ada Lovelace: job_id 103
 - Margaret Hamilton: job_id 103
- **department_id 40, job_id 106:** Có 1 nhân viên.
 - Elon Musk: job_id 106

Tổng kết

Câu lệnh `SELECT department_id, job_id, COUNT(*) FROM employees GROUP BY department_id, job_id;` giúp bạn đếm số lượng nhân viên cho mỗi cặp `department_id` và `job_id`. Câu lệnh này sử dụng `GROUP BY department_id, job_id` để nhóm các hàng theo cả hai cột `department_id` và `job_id`, và áp dụng hàm tổng hợp `COUNT` để đếm số lượng hàng trong từng nhóm.

| employee_id | first_name | last_name | department_id | job_id | salary |
|-------------|------------|-----------|---------------|---------|--------|
| 1 | John | Doe | 10 | IT_PROG | 60000 |
| 2 | Jane | Smith | 20 | IT_PROG | 75000 |
| 3 | Larry | Page | 10 | IT_PROG | 65000 |
| 4 | Sergey | Brin | 30 | HR | 80000 |
| 5 | Alan | Turing | 20 | IT_PROG | 70000 |
| 6 | Ada | Lovelace | 30 | IT_PROG | 90000 |
| 7 | Charles | Babbage | 10 | IT_PROG | 72000 |
| 8 | Grace | Hopper | 20 | HR | 80000 |
| 9 | Margaret | Hamilton | 30 | IT_PROG | 85000 |
| 10 | Elon | Musk | 40 | IT_PROG | 95000 |

▼ Tính lương trung bình (`avg_salary`) cho mỗi `department_id` của các nhân viên có `job_id` là 'IT_PROG'

```
SELECT department_id, AVG(salary) AS avg_salary
FROM employees
WHERE job_id = 'IT_PROG'
GROUP BY department_id;
```

Kết quả

Kết quả của câu lệnh sẽ như sau:

| department_id | avg_salary |
|---------------|------------|
| 10 | 65666.67 |
| 20 | 71666.67 |
| 30 | 87500 |
| 40 | 95000 |

Giải thích kết quả

- **department_id 10:** Có 3 nhân viên với `job_id` là 'IT_PROG' và lương trung bình là 65666.67.
 - John Doe: 60000
 - Larry Page: 65000
 - Charles Babbage: 72000
- **department_id 20:** Có 2 nhân viên với `job_id` là 'IT_PROG' và lương trung bình là 71666.67.
 - Jane Smith: 75000
 - Alan Turing: 70000
- **department_id 30:** Có 2 nhân viên với `job_id` là 'IT_PROG' và lương trung bình là 87500.
 - Ada Lovelace: 90000
 - Margaret Hamilton: 85000
- **department_id 40:** Có 1 nhân viên với `job_id` là 'IT_PROG' và lương trung bình là 95000.
 - Elon Musk: 95000

Tổng kết

Câu lệnh `SELECT department_id, AVG(salary) AS avg_salary FROM employees WHERE job_id = 'IT_PROG' GROUP BY department_id;` giúp bạn tính lương trung bình cho mỗi `department_id` của các nhân viên có `job_id` là 'IT_PROG'. Câu lệnh này sử dụng `WHERE job_id = 'IT_PROG'` để lọc các nhân viên có `job_id` là 'IT_PROG', sau đó nhóm các hàng theo `department_id` bằng `GROUP BY department_id` và áp dụng hàm tổng hợp `AVG` để tính toán lương trung bình trong từng nhóm.

| employee_id | first_name | last_name | department_id | job_id | salary |
|-------------|------------|-----------|---------------|---------|--------|
| 1 | John | Doe | 10 | IT_PROG | 60000 |
| 2 | Jane | Smith | 20 | IT_PROG | 75000 |
| 3 | Larry | Page | 10 | IT_PROG | 65000 |
| 4 | Sergey | Brin | 30 | HR | 80000 |
| 5 | Alan | Turing | 20 | IT_PROG | 70000 |

| | | | | | |
|----|----------|----------|----|---------|-------|
| 6 | Ada | Lovelace | 30 | IT_PROG | 90000 |
| 7 | Charles | Babbage | 10 | IT_PROG | 72000 |
| 8 | Grace | Hopper | 20 | HR | 80000 |
| 9 | Margaret | Hamilton | 30 | IT_PROG | 85000 |
| 10 | Elon | Musk | 40 | IT_PROG | 95000 |

▼ Sử dụng **HAVING** nhằm lọc kết quả sau khi nhóm, ví dụ này sẽ chỉ trả về các **department_id** có lương trung bình (**avg_salary**) lớn hơn 50,000.

```
SELECT department_id, AVG(salary) AS avg_salary
FROM employees
GROUP BY department_id
HAVING AVG(salary) > 50000;
```

Kết quả

Kết quả của câu lệnh sẽ như sau:

| department_id | avg_salary |
|---------------|------------|
| 10 | 65666.67 |
| 20 | 75000 |
| 30 | 85000 |
| 40 | 95000 |

Giải thích kết quả

- **department_id 10:** Lương trung bình là 65666.67.
 - John Doe: 60000
 - Larry Page: 65000
 - Charles Babbage: 72000
- **department_id 20:** Lương trung bình là 75000.
 - Jane Smith: 75000
 - Alan Turing: 70000
 - Grace Hopper: 80000
- **department_id 30:** Lương trung bình là 85000.
 - Sergey Brin: 80000
 - Ada Lovelace: 90000
 - Margaret Hamilton: 85000

- **department_id 40:** Lương trung bình là 95000.
 - Elon Musk: 95000

Các `department_id` này đều có lương trung bình lớn hơn 50,000, nên tất cả đều được trả về.

Tổng kết

Câu lệnh `SELECT department_id, AVG(salary) AS avg_salary FROM employees GROUP BY department_id HAVING AVG(salary) > 50000;` giúp bạn tính lương trung bình cho mỗi `department_id` và chỉ trả về các phòng ban có lương trung bình lớn hơn 50,000. Câu lệnh này sử dụng `GROUP BY department_id` để nhóm các hàng theo `department_id`, `AVG(salary)` để tính lương trung bình, và `HAVING AVG(salary) > 50000` để lọc các kết quả sau khi nhóm.

▼ LƯU Ý

Lưu ý:

- Mệnh đề `GROUP BY` phải được sử dụng cùng với các hàm tổng hợp như `COUNT`, `SUM`, `AVG`, `MIN`, `MAX`.
- Các cột được sử dụng trong mệnh đề `GROUP BY` phải được có trong mệnh đề `SELECT`.
- Mệnh đề `HAVING` được sử dụng để lọc các nhóm kết quả sau khi đã được nhóm theo `GROUP BY`.
- `HAVING` được sử dụng để lọc các nhóm sau khi đã áp dụng `GROUP BY`, trong khi `WHERE` được sử dụng để lọc các hàng trước khi áp dụng `GROUP BY`.
- Sử dụng `GROUP BY` một cách cẩn thận, vì nó có thể ảnh hưởng đến hiệu suất truy vấn, đặc biệt là khi nhóm dữ liệu lớn.

Ngoài ra, **GROUP BY** còn có một số cách sử dụng khác:

- Tạo các báo cáo tóm tắt dữ liệu.
- Xác định các giá trị duy nhất trong một cột.
- Kết hợp dữ liệu từ nhiều bảng.

Bảng dữ liệu mẫu: `employees`

Giả sử chúng ta có bảng `employees` với dữ liệu sau:

| employee_id | first_name | last_name | department_id | salary |
|-------------|------------|-----------|---------------|--------|
| 1 | John | Doe | 10 | 60000 |
| 2 | Jane | Smith | 20 | 75000 |
| 3 | Larry | Page | 10 | 65000 |
| 4 | Sergey | Brin | 30 | 80000 |
| 5 | Alan | Turing | 20 | 70000 |
| 6 | Ada | Lovelace | 30 | 90000 |
| 7 | Charles | Babbage | 10 | 72000 |
| 8 | Grace | Hopper | 20 | 80000 |
| 9 | Margaret | Hamilton | 30 | 85000 |

| | | | | |
|----|------|------|----|-------|
| 10 | Elon | Musk | 40 | 95000 |
|----|------|------|----|-------|

▼ **Đếm số lượng nhân viên trong mỗi phòng ban (COUNT)**

```
SELECT department_id, COUNT(*) AS num_employees
FROM employees
GROUP BY department_id;
```

Kết quả:

| department_id | num_employees |
|---------------|---------------|
| 10 | 3 |
| 20 | 3 |
| 30 | 3 |
| 40 | 1 |

▼ **Tính tổng lương trong mỗi phòng ban (SUM)**

```
SELECT department_id, SUM(salary) AS total_salary
FROM employees
GROUP BY department_id;
```

Kết quả:

| department_id | total_salary |
|---------------|--------------|
| 10 | 197000 |
| 20 | 225000 |
| 30 | 255000 |
| 40 | 95000 |

▼ **Tính lương trung bình trong mỗi phòng ban (AVG)**

```
SELECT department_id, MAX(salary) AS max_salary
FROM employees
GROUP BY department_id;
```

Kết quả:

| department_id | max_salary |
|---------------|------------|
| 10 | 72000 |
| 20 | 80000 |
| 30 | 90000 |
| 40 | 95000 |

▼ **Tìm lương thấp nhất trong mỗi phòng ban (MIN)**

```
SELECT department_id, MIN(salary) AS min_salary
FROM employees
GROUP BY department_id;
```

Kết quả:

| department_id | min_salary |
|---------------|------------|
| 10 | 60000 |
| 20 | 70000 |
| 30 | 80000 |
| 40 | 95000 |

Giải thích câu lệnh SQL

- **SELECT department_id, COUNT(*) AS num_employees:** Chọn cột `department_id` và đếm số lượng nhân viên trong mỗi phòng ban.
- **GROUP BY department_id:** Nhóm các hàng theo `department_id`.
- **AS:** Đặt tên cho cột kết quả.

Các hàm tổng hợp khác (`SUM`, `AVG`, `MAX`, `MIN`) hoạt động tương tự, nhưng thay vì đếm số lượng, chúng tính toán tổng, trung bình, giá trị lớn nhất, và giá trị nhỏ nhất tương ứng.