Partition by / Group by

Partition by ->>>>>>>>>Result set is return each group <<<<<<<-

You need to identify the group.

A group can be a column , or multiple columns.

Work with window function:

avg() , sum() , count(\*)

rank(), row\_number() : These two are similar . But when two persons have same salary , rank() will return same number , but row\_number will return different number.

用Partition by唔會影響到原生table 行數,會多左一行alias column as result.

For example:

A black table with white text

AI-generated content may be incorrect.

7 Tom Sales 60000 2017-03-19

***Example: To rank salesman in each department order by salary desc.***

rank() vs row\_number()

Select id, name, department, salary,

rank() over (partition by department order by salary desc) AS salary\_rank

From table;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | name | department | salary | salary\_rank |
| 4 | David | IT | 90000 | 1 |
| 2 | Bob | IT | 80000 | 2 |
| 3 | Charile | HR | 75000 | 1 |
| 1 | Alice | HR | 70000 | 2 |
| 7 | Tom | Sales | 60000 | 1 |
| 6 | Frank | Sales | 60000 | 1 |
| 5 | Eve | Sales | 50000 | 2 |

Select id, name, department, salary,

Row\_number() over (partition by department order by salary desc) AS salary\_rank

From table;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | name | department | salary | salary\_rank |
| 4 | David | IT | 90000 | 1 |
| 2 | Bob | IT | 80000 | 2 |
| 3 | Charile | HR | 75000 | 1 |
| 1 | Alice | HR | 70000 | 2 |
| 7 | Tom | Sales | 60000 | 1 |
| 6 | Frank | Sales | 60000 | 2 |
| 5 | Eve | Sales | 50000 | 2 |

Partition by can -> order by multiple columns:

Select id,name,department,salary,hire\_date,

avg(salary) over (Partition by department order by salary desc, hire\_date) as avg\_salary

from table;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | name | department | salary | hire\_date | sum\_salary |
| 4 | David | IT | 90000 | 2019-11-02 | 85,000 |
| 2 | Bob | IT | 80000 | 2020-03-22 | 85,000 |
| 3 | Charile | HR | 75000 | 2022-08-10 | 72,500 |
| 1 | Alice | HR | 70000 | 2021-05-15 | 72,500 |
| 7 | Tom | Sales | 60000 | 2017-03-19 | 56,666.666 |
| 6 | Frank | Sales | 60000 | 2022-12-05 | 56,666.666 |
| 5 | Eve | Sales | 50000 | 2023-01-20 | 56,666.666 |

In each partition, Order by salary first , then order by hire\_date in default ASC.

Partition by can -> partition by multiple columns:

Select id,name,department,salary,hire\_date,

avg(salary) over (Partition by department,name ) as avg\_salary

from table;

What happen if :

SELECT  
id,  
salesperson,  
amount,  
sale\_date,  
SUM(amount) OVER (PARTITION BY salesperson ORDER BY sale\_date) AS cumulative\_sales  
FROM sales order by cumulative\_sales;

cumulative\_sales is not recognized in the ORDER BY clause because it was defined in the SELECT clause

Use subquery !!

Can you tell the sequences flow of how sql engine process the sql:

1. From table, scan the table
2. Where condition
3. Select
4. Order by

(However , order by )

SELECT \* FROM ( SELECT id, salesperson, amount, sale\_date, SUM(amount) OVER (PARTITION BY salesperson ORDER BY sale\_date) AS cumulative\_sales FROM sales ) AS subquery ORDER BY cumulative\_sales;