Grouping Data

GROUP BY statement groups rows with the same values in a particular column. The GROUP BY clause is also used along with aggregate functions to perform various actions on the database.

The GROUP BY statement groups rows that have the same values into summary rows, like "find the number of customers in each country". The GROUP BY statement is often used with aggregate functions to group the result-set by one or more columns.

GROUP BY SYNTAX:

```
SELECT column_name(s)
FROM table_name
WHERE condition
GROUP BY column_name(s)
ORDER BY column_name(s);
```

GROUP BY and DISTINCT statements are used to find the distinct elements in the table.

We can find the distinct element in the table using two ways

1. SELECT column_name FROM table_name GROUP BY column_name;
2. SELECT DISTINCT(column_name) FROM table_name;

Both queries will have the same output.

ALIASES

SQL aliases are used to give a table, or a column in a table, a temporary name. Aliases are often used to make column names more readable. An alias only exists for the duration of that query. An alias is created with the AS keyword.

General Syntax:

Alias Column

```
SELECT column_name AS alias_name
FROM table_name;
```

Alias Table

```
SELECT column_name(s)
FROM table_name AS alias_name;
```

Aggregate Functions

Functions that take multiple values as input, perform calculations on them, return a single value as output is known as aggregate functions. Aggregate functions are often used along with the GROUP BY clause to perform calculations

1. COUNT()
2. SUM()
3. AVG()
4. MIN()
5. MAX()

HAVING

The HAVING clause was added to SQL because the WHERE keyword cannot be used with aggregate functions.
General Syntax:

```
SELECT column_name(s)
FROM table_name
WHERE condition
GROUP BY column_name(s)
HAVING condition
ORDER BY column_name(s);
```

HAVING and WHERE clauses have the same function of filtering the row value in SQL. WHERE clause is used when there is a need to filter the records from a table that is based on a specified condition, Having Clause is used to filter the record from the groups based on the specified condition. Group by clause is necessary while using the HAVING clause. HAVING is used after the GROUP BY clause while WHERE is used before the GROUP BY clause.

```
Q: Calculate the total sales for each client?
SELECT CLIENT_ID, SUM(INVOICE_TOTAL) AS TOTAL_SALES
FROM INVOICES
GROUP BY CLIENT_ID;
```

```
Q: Show the clients who have total sales more than 5000?
```

```
I)
SELECT CLIENT_ID, SUM(INVOICE_TOTAL) AS TOTAL_SALES
FROM INVOICES
WHERE TOTAL_SALES > 5000
GROUP BY CLIENT_ID;
```

```
II)
SELECT CLIENT_ID, SUM(INVOICE_TOTAL) AS TOTAL_SALES
FROM INVOICES GROUP BY CLIENT_ID
HAVING TOTAL_SALES > 5000;
```

Here, first solution is wrong because we can not use total sales before grouping up. We can use total sales after group by only. So, second solution is correct.

In summary, WHERE is used to filter rows before aggregation or grouping, while HAVING is used to filter grouped rows after aggregation.