GROUPING DATA WITH GROUP BY

GROUP BY clause is used to group or categorize the data. In other words it divide rows in a table into smaller groups. We can then use the group functions to return summary information for each group.

If no GROUP BY clause is specified, then the default grouping is the entire result set. When the query executes and the data is fetched, it is grouped based on the GROUP BY clause and the group function is applied.

Syntax of GROUP BY

```
SELECT column,group_function(column) FROM table
[WHERE condition]
[GROUP BY   group_by_expression]
[ORDER BY column];
```

Here, group_by_expression specifies columns whose values determine the basis for grouping rows.

For example, If we have to find the total salary of each department manually, first we group the records on the basis of department number and then we apply the sum function on salary of each group to obtain the required result.

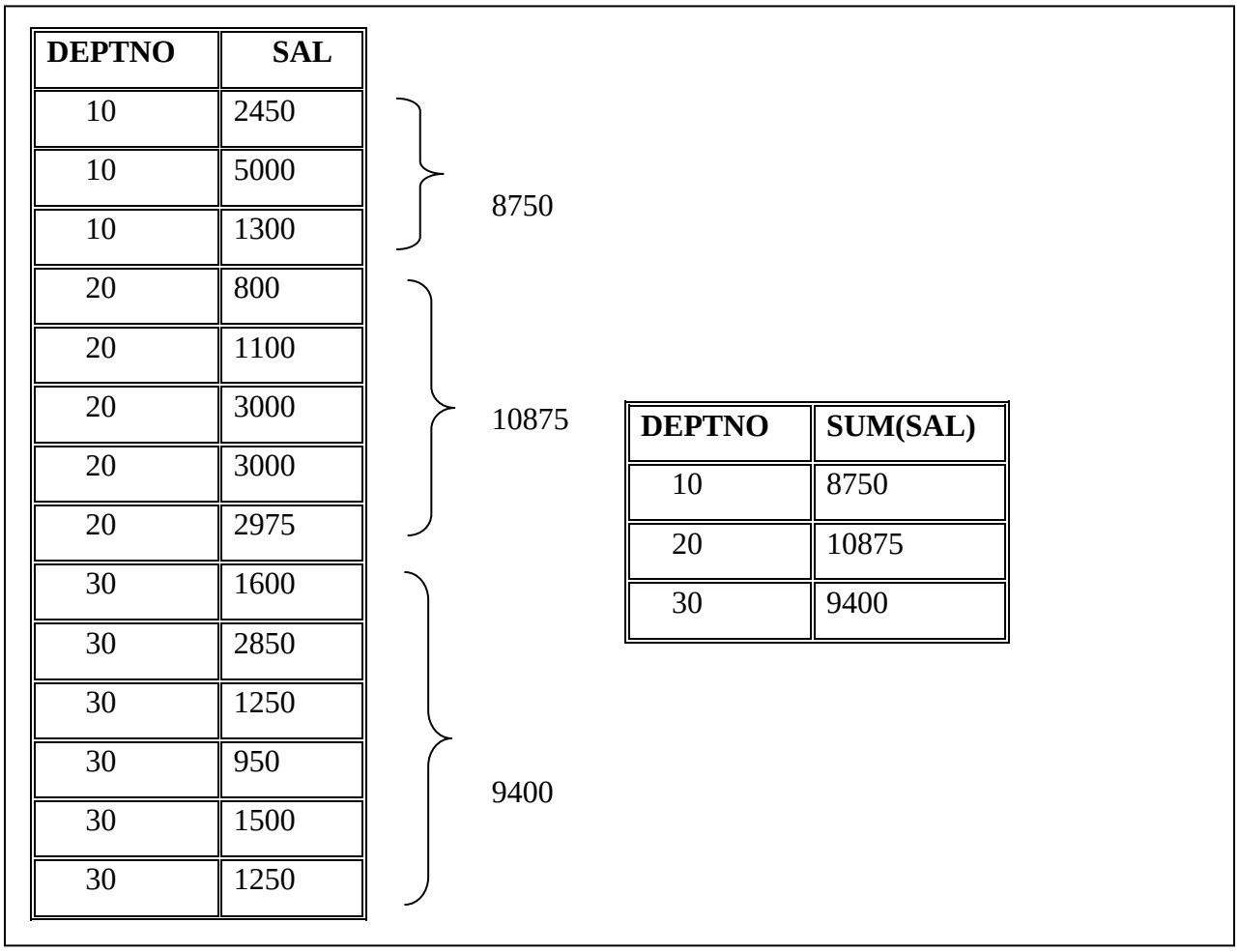
Similarly in SQL we apply the GROUP BY clause on deptno and then calculate the total salary for each group by Sum(sal) function as shown below:

SQL>*SELECT deptno, Sum(sal) FROM emp GROUP BY deptno;*

OUTPUT:

DEPTNO	SUM(SAL)
10	2916.6667
20	2175
30	1566.6667

The below figure shows the grouping and execution of query:



DEPTNO	SAL
10	2450
10	5000
10	1300
20	800
20	1100
20	3000
20	3000
20	2975
30	1600
30	2850
30	1250
30	950
30	1500
30	1250

DEPTNO	SUM(SAL)
10	8750
20	10875
30	9400

Here is how this SELECT statement, containing a GROUP BY clause, is evaluated:

- The SELECT clause specifies the columns to be retrieved i.e Department number column in the EMP table, the sum of all the salaries in the group you specified in the GROUP By clause
- The FROM clause specifies the tables that the database must access i.e EMP table.
- The WHERE clause specifies the rows to be retrieved. Since there is no WHERE clause, by default all rows are retrieved.

The GROUP BY clause specifies how the rows should be grouped. Department number groups the rows, so the AVG function that is being applied to the salary column will calculate the average salary for each department.

- List the average salary of each job in the emp table.

SQL>*SELECT JOB,AVG(SAL) FROM EMP GROUP BY JOB;*

OUTPUT:

JOB	AVG(SAL)
ANALYST	3166.5
CLERK	962.5
MANAGER	2758.3333
PRESIDENT	5000
SALESMAN	1400

- List the maximum salary for each dept.

SQL>*SELECT DEPTNO,MAX(SAL) FROM EMP GROUP BY DEPTNO;*

OUTPUT:

DEPTNO	MAX(SAL)
10	5000
20	3333
30	2850