**Chapter 4**

**Columns and Groupings**

**Objectives**

* Implement GROUP BY modifier on columns;
* Work on aggregate functions;
* Filter groupings using HAVING clause;
* Use DISTINCT to eliminate row duplication.

Most of our scripts that we have done so far are in details. Actual databases consist of big tables with thousands (or even million) of rows. But there are requirements which do not need details but summary of the content. We may use GROUP BY modifier in parallel with AGGREGATE functions to compute groupings in column.

GROUP BY syntax is as follows:

SELECT <column name 1>, … <column name n> FROM <employee>

GROUP BY < column groupings >

Connect again to SAMPLE database so we may work with examples.

Using the EMPLOYEE table, list the JOB’s positions held within the company:

SELECT job FROM employee

GROUP BY job

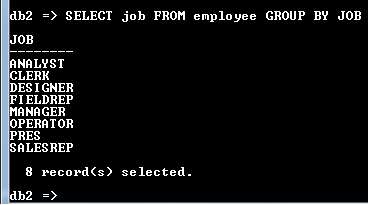


Figure 4-1. Using GROUP BY

The EMPLOYEE’s table consists of 42 employees. These employees hold various positions in the company. The SELECT statement GROUPs job positions and lists it in alphabetical order.

So, if we want to list all the departments of those employees working at:

SELECT workDept FROM employee

GROUP BY workDept

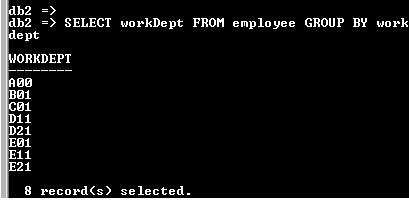


Figure 4-2. Another GROUP BY example

**Use Aggregate functions**

Aggregate functions are used to summarize details either by counting, summing up, averaging, getting its minimum, or its maximum value. The following functions may be associated with GROUP BY modifier:

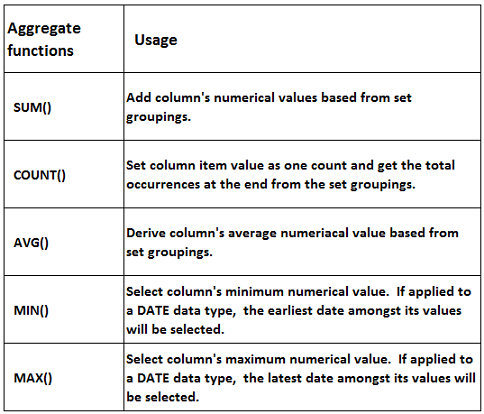


Table 4-1. Aggregate functions

Now, if we become more curious about the company we have in our database, we may inquire how many employees occupy a particular position. Here, we associate GROUP BY modifier with COUNT() aggregate function.

SELECT job, COUNT(job) FROM employee

GROUP BY job

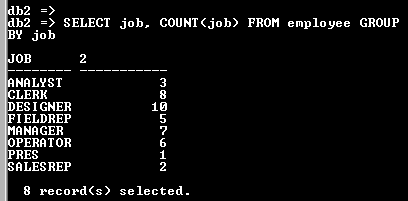


Figure 4-3. COUNT() example

This time we have listed two columns: job and COUNT(job). The GROUP BY modifier repeats the summary of job positions and the second column serves as a COUNT of those positions. The 2 in the column header represents the column title as it is said to be the second column. We may improve our listings by assigning an alias name.

SELECT job, COUNT(job) “Employee Count” FROM employee

GROUP BY job

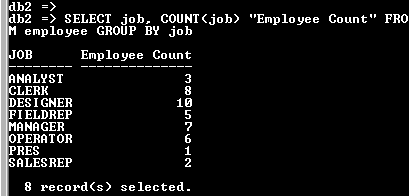


Figure 4-4. Assigning alias name with COUNT() function

COUNT() function is best served in tallying row occurences of non-numerical value. This is the only aggregate function that can be used in text data type columns.

Although employees may hold same position within the company, still their salary differs.

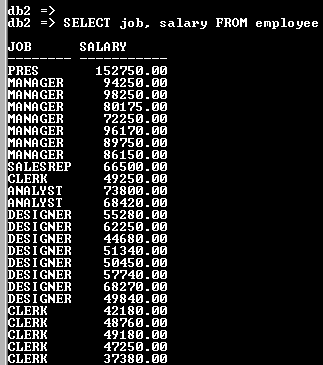


Figure 4-5. With same position but different salary

Our next line of inquiry is to ask what would be the average salary for each position in the company. Execute the given script:

SELECT job, AVG(salary) “Average Salary” FROM employee

GROUP BY job

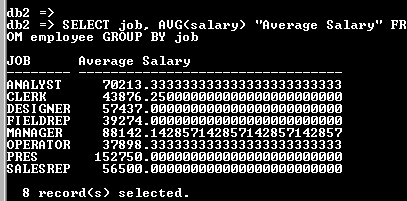


Figure 4-6. Using AVG() function

What would be the SUM of the annual salaries for each position if budget allocation is required? We have:

SELECT job, SUM(salary) “Sum Paid Annually” FROM employee

GROUP BY job

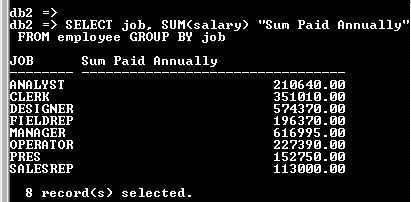


Figure 4-7. Using SUM() function

And if we want to check what is the MINimum salary paid for each position:

SELECT job, MIN(salary) “Minimum Salary” FROM employee

GROUP BY job

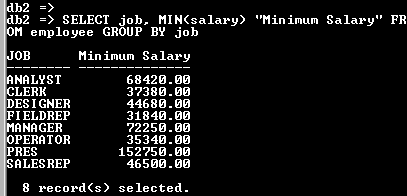


Figure 4-8. Using MIN() function

Getting the MAXimum salary for the position is almost the same as we do for the minimum salary:

SELECT job, MAX(salary) “Maximum Salary” FROM employee

GROUP BY job

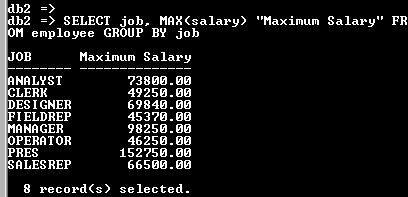


Figure 4-9. Using MAX() function

MIN() and MAX() functions are not limited to numerical values. It may also be applied to DATE data types. We inquire what is the earliest hiring date occurred within the company, we have:

SELECT MIN(hireDate) “Earliest Hiring Date” FROM employee



Figure 4-10. Using MIN() on DATE data type

To get the latest hiring date:

SELECT MAX(hireDate) “Latest Hiring Date” FROM employee

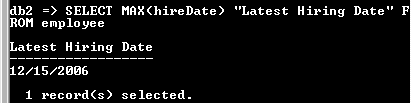


Figure 4-11. Using MAX() on DATE data type

Notice that we do not use GROUP BY from the two previous examples. Our focus is on the entire company that is why it is unnecessary to group. You may use this also on other aggregate functions depending on the requirements set.

Get the total annual salary of the company:

SELECT SUM(salary) “Total Annual Salary” FROM employee



Figure 4-12. Using SUM() without GROUP BY

**Filter GROUP BY using HAVING clause**

You can still select rows on the output of GROUP BY modifier. Use HAVING clause together with the condition you have defined:

SELECT <column name 1>, … <column name n> FROM <table name>

GROUP BY <group of columns> HAVING <set of conditions>

Think of HAVING as a WHERE filter when you want to isolate result of a GROUP BY modifier. When we say , HAVING with set of conditions, the clause is not limited with one condition but may work with compound or multiple conditions just as you work with WHERE filter conditions.

If we want to isolate departments A00, D11, and E11 to track how many employees at EMPLOYEE table of SAMPLE datatabase, we have:

SELECT workDept, COUNT(empno) “Total Employees” FROM employee

GROUP BY workDept HAVING workDept IN (‘A00’, ‘D11’, ‘E11’)

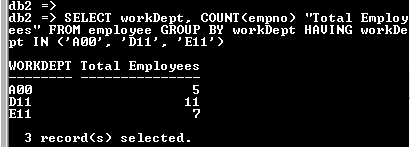


Figure 4-13. Using HAVING clause

Our requirement needs to count the number of employees who occupies the DESIGNER and ANALYST job positions:

SELECT job, COUNT(job) “Employees Count” FROM employee

GROUP BY job HAVING job =’ANALYST’ OR job=’DESIGNER’

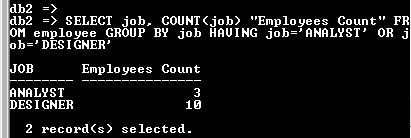


Figure 4-14. HAVING with compound conditions

Instead of IN clause, we use from the previous example logical OR to select ANALYST and DESIGNER positions.

**Group with multiple columns**

GROUP BY does not work only with one column. You may set multiple columns depending on the requirement. What we have discussed only focuses with one column. Let us have examples with multiple columns.

The need is to list all the MALE and FEMALE per each department:

SELECT workDept, sex, COUNT(sex) “Employees Count” FROM employee

GROUP BY workDept, sex

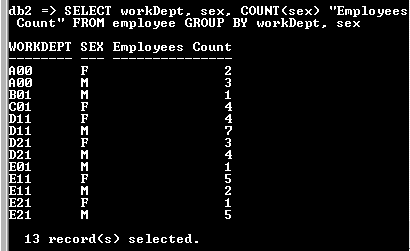


Figure 4-15. Working with multiple columns

The SELECT statement here directs to make two sets of grouping: one with workDept and another by sex. The grouping will work with the outer most column (sex) first before proceeding with the inner most column (workDept).

If we want to report the number of positions hired per job per department and by hiring year, we have:

SELECT YEAR(hireDate) “Year Hired”, workDept, job, COUNT(empno) “Employees Count”

FROM employee

GROUP BY YEAR(hireDate), workDept, job



Figure 4-16. Working with 3 sets of grouping

**Work with WHERE, GROUP BY, HAVING, ORDER BY clauses in one statement**

Though it is seldom you have a situation that requires to have WHERE, GROUP BY, HAVING, and ORDER BY clauses in a single SELECT statement, you may execute it as long as it is in proper order. SQL is arranged in intuitive manner patterned to an english language. The key here is to make your script comprehensible as stated in an english statement. Chances that it will work are high if you state it right and the statement sounds clear.

SELECT <column name 1>, … <column name n> FROM <table name>

WHERE <set of conditions>

GROUP BY <group of columns>

HAVING <set of conditions>

ORDER BY <group of columns> [ASC/DESC]

Our next requirement is to display employees’ maximum salary per department isolating those who have education equal or higher than 16 years. Furthermore, we want maximum salary ranging from 60,000 up to 160,000 dollars annually. Sort the result according to department and years of education (highest education is listed first).

SELECT workDept, edLevel, MAX(salary) “Maximum Salary” FROM employee

WHERE edLevel >= 16

GROUP BY workDept, edLevel

HAVING MAX(salary) BETWEEN 60000 AND 160000

ORDER BY workDept, edLevel DESC

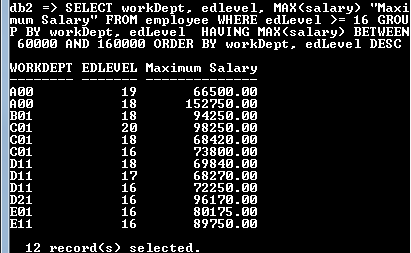


Figure 4-17. SELECT complex statement

**Use DISTINCT to eliminate multiple row duplication**

How do we know the departments of EMPLOYEE table of SAMPLE database without using GROUP BY modifier? As an alternative, employ the use of DISTINCT.

SELECT DISTINCT workdept FROM employee

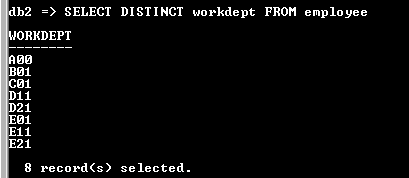


Figure 4-18. Using DISTINCTon one column

It may also work with multiple columns. Let us display the job positions occupied in a department.

SELECT DISTINCT workDept, job FROM employee

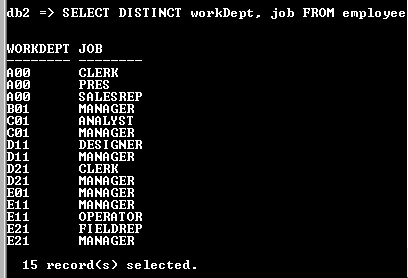


Figure 4-19. DISTINCT with multiple rows

**Chapter Exercises**

I. **Multiple Choice:**  Refer to SAMPLE database. Select the letter of the correct answer.

1. How do you get the total employees count of EMPLOYEE table?
2. SELECT empNo, COUNT(empNo) “Employees Count” FROM employee GROUP BY empNo
3. SELECT COUNT(empNo) “Employees Count” FROM employee
4. SELECT lastName, SUM(lastName) “Employees Count” FROM employee GROUP BY lastName
5. SELECT SUM(lastName) “Employees Count” FROM employee
6. Which SQL script will get the total number of female and male of the company?
7. SELECT COUNT(sex) “Total” FROM employee
8. SELECT SUM(sex) “Total” FROM employee
9. SELECT sex, COUNT(sex) “Total” FROM employee GROUP BY sex
10. SELECT sex, SUM(sex) “Total” FROM employee GROUP BY sex
11. What is the number of employees per department?
12. SELECT workDept, COUNT(empNo) FROM employee GROUP BY workDept, empNo
13. SELECT workDept, SUM(workDept) FROM employee GROUP BY workDept, empNo
14. SELECT workDept, COUNT(empno) FROM employee GROUP BY workDept
15. SELECT workDept, SUM(workDept) FROM employee GROUP BY workDept
16. What is the simpliest equivalent statement for:

SELECT job FROM employee

GROUP BY job

1. SELECT DISTINCT job FROM employee
2. SELECT job FROM employee
3. SELECT DISTINCT job FROM employee GROUP BY job
4. SELECT job FROM employee ORDER BY job
5. How will you group employees based on month hired (regardless of the year)?
6. SELECT hireDate, lastName, firstNme FROM employee GROUP BY hireDate
7. SELECT MONTH(hireDate) , lastName, firstNme FROM employee GROUP BY MONTH(hireDate)
8. SELECT DISTINCT MONTH(hireDate), lastName, firstNme FROM employee
9. SELECT YEAR(hireDate), lastName, firstNme FROM employee GROUP BY MONTH(hireDate)
10. If you are asked to total employees according to years of education, which statement it is?
11. SELECT edLevel FROM employee GROUP BY edLevel
12. SELECT DISTINCT COUNT(edLevel) FROM employee
13. SELECT edLevel, COUNT(edLevel) FROM employee
14. SELECT edLevel, COUNT(edLevel) FROM employee GROUP BY edLevel
15. The company wants to see the average salary difference between male and female by job position. Which of the following will do the trick?
16. SELECT job, sex, AVG(salary) “Average Salary” FROM employee
17. SELECT DISTINCT job, sex, AVG(salary) “Average Salary” FROM employee
18. SELECT job, sex, AVG(salary) “Average Salary” FROM employee GROUP BY sex
19. SELECT job, sex, AVG(salary) “Average Salary” FROM employee GROUP BY job, sex
20. Count the number of employees per department. List only departments which have 7 and up employees. Which script will do this?
21. SELECT workDept, COUNT(empNo) “Number of Employees” FROM employee
22. SELECT workDept, COUNT(empNo) “Number of Employees” FROM employee GROUP BY workDept
23. SELECT workDept, COUNT(empNo) “Number of Employees” FROM employee GROUP BY workDept HAVING COUNT(empNo) > 6
24. SELECT workDept, SUM(empNo) “Number of Employees” FROM employee GROUP BY workDept HAVING SUM(empNo) >= 7
25. How will you list all job positions in the company held by females. Display how many is holding that position by department and what is the average salary (highest salary first). Include only salary 50,000 up.
26. SELECT workDept, job, sex, COUNT(sex) “EmpCount”, AVG(salary) “AvgSal” FROM employee

WHERE sex = ‘F’

GROUP BY workDept, job, sex

HAVING AVG(salary) >= 50000

ORDER BY 5 DESC

1. SELECT workDept, job, sex, COUNT(sex) “EmpCount”, AVG(salary) “AvgSal” FROM employee

WHERE sex <> ‘M’

GROUP BY workDept, job, sex

HAVING salary > 50000

ORDER BY AVG(salary) DESC

1. SELECT workDept, job, sex, COUNT(sex) “EmpCount”, salary “AvgSal” FROM employee

WHERE sex = ‘F’

GROUP BY workDept, job, sex

HAVING AVG(salary) >= 50000

ORDER BY AVG(salary) ASC

1. SELECT workDept, job, sex, COUNT(sex) “EmpCount”, AVG(salary) “AvgSal” FROM employee

WHERE sex = ‘F’

GROUP BY workDept, job, sex, COUNT(sex)

HAVING AVG(salary) >= 50000

ORDER BY 5

1. What is the minimum and maximum salary per job position by department? Display also the number of employees holding specific position. List according to department and by highest salary first.
2. SELECT workDept, job, COUNT(empNo) “Emp Count”, MIN(salary) “Minimum Salary”, MAX(salary) “Maximum Salary FROM employee

GROUP BY workDept, job

ORDER BY 5 DESC

1. SELECT workDept, job, empNo “Emp Count”, MIN(salary) “Minimum Salary”, MAX(salary) “Maximum Salary FROM employee

GROUP BY workDept, job

1. SELECT workDept, job, COUNT(empNo) “Emp Count”, MIN(salary) “Minimum Salary”, MAX(salary) “Maximum Salary FROM employee

GROUP BY workDept, job

ORDER BY 1, 5 DESC

1. SELECT workDept, job, COUNT(empNo) “Emp Count”, MIN(salary) “Minimum Salary”, MAX(salary) “Maximum Salary FROM employee

GROUP BY workDept, job

ORDER BY workDept, MAX(salary)

II. **Problems**

Refer to HOPE database. If you have not created the database (or prefer a new one), click this link to download and install all tables under the database. Make sure to SET SCHEMA store after you CONNECT TO hope to access the tables.

1. Determine the number of customers who pays cash-on-delivery, 30-day, and 45-day notice.
2. Create number of customers within California that pays cash-on-delivery, 30-day, and 45-day notice.
3. How many current employees does the company has? Exclude those who are already separated.
4. How many current male and female does the company has?
5. The company wants to monitor repeat customers. Display customer number and the number of transactions the customer has entered to the company. List with highest transactions first. Use SALES table.
6. Which transaction has the biggest number of total quantity delivered? Refer to the SALESDETAIL table. Sort from the highest to the lowest quantity.

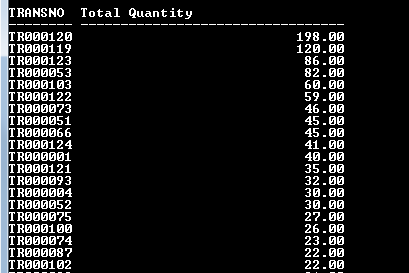


Figure 4-20. Partial output of SALESDETAIL table

1. Transaction is sometimes paid in multiple payments. Which transaction paid the biggest amount for the year 2010? Refer to PAYMENT table.
2. Generate the total payment made per transaction by year. Exclude total amount less than 1,000 dollars. List according to year and highest amount first.



Figure 4-21. Partial list of PAYMENT transaction

1. What are the units used in PRODUCT table? Do not use GROUP BY to display result.
2. Which product code which has multiple changes in prices in PRICEHIST table?

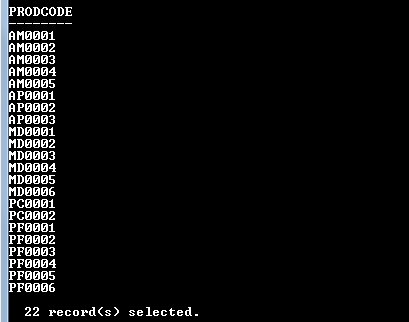


Figure 4-22. Product codes that fluctuate unit price more than once

**Answers to selected numbers**

I. Multiple choice

1. B

3. C

5. B

7. D

9. A

II. Problems

1. SELECT payTerms, COUNT(custNo) “Number of Customer” FROM customer

GROUP BY payTerms

1. SELECT COUNT(empNo) “Current Employees” FROM employee

WHERE sepDate IS NULL

1. SELECT custNo, COUNT(transNo) “Number of Transactions” FROM sales

GROUP BY custNo

ORDER BY COUNT(transNo) DESC

1. SELECT transNo, SUM(amount) “Pay amount” FROM salesDetail

WHERE YEAR(payDate) = 2010

GROUP BY transNo

ORDER BY 2 DESC

1. SELECT DISTINCT(unit) FROM product