5

Reporting Aggregate Data

Using the Group Functions

Objectives

After completing this lesson AND

practicing, you will begin to understand the following:

• Describe the use of group functions

• Group data by using the GROUP BY clause

• Include or exclude grouped rows by using the

HAVING Clause

This lesson further addresses functions.

It focuses on obtaining summary information (such as averages) for groups of rows.

It discusses how to group rows in a table into smaller sets and

* how to specify search criteria for groups of rows.

Lesson Agenda

Group Functions

Grouping Rows

Nesting Group Functions

What Are Group Functions

Group Functions 🡺 operate on sets of rows

🡺 to give one result per group of rows

PROBLEM:

Find the average salary for all employees.

EXAMPLE:

SELECT AVG (salary)

FROM employee;

AVG(SALARY)

One result from a set of rows.

In this case the whole table

8775

You can do a lot of reporting using single group functions

PROBLEM:

**Find the**

**1 average salary,**

**2 minimum salary,**

**3 the difference**

**4 the highest salary**

**5 and the difference.**

=========================

SELECT AVG(SALARY),

MIN(SALARY),

AVG(SALARY) - MIN(SALARY) as "MIN. DIFF",

MAX(SALARY),

AVG(SALARY)- MAX(SALARY) as "MAX DIFF"

FROM EMPLOYEE;

AVG(SALARY) MIN(SALARY) MIN. DIFF MAX(SALARY) MAX DIFF

----------- ----------- ---------- ----------- ----------

8725 2500 6225 24000 -15275

**Single Row functions** worked on single rows and returned 1 result per row

SELECT UPPER(last\_name) …

Each row selected changed the format of whatever last\_name was stored as, to display in UPPER case

**Group functions (multi-row functions)**

🡪 Operate on sets of rows to give one result per group.

- These sets may comprise the entire table or the table split into groups

**Types of Group Functions**

AVG

COUNT

MAX

Group

Functions

MIN

1 result

STDDEV

SUM

VARIANCE

Each function acts on a set of rows determined by the SELECT conditions

GROUP FUNCTION SYNTAX

PROBLEM:

Start with this problem 🡪

President wants to know data about salaries, such as AVERAGE, what the highest and lowest paid person's salary is and the company's total salary payout.

🡺 🡺 GO BACK to previous page and find the appropriate function

**SELECT AVG (SALARY),**

**MAX (SALARY), -- highest paid**

**MIN (SALARY), -- lowest paid**

**SUM (SALARY) -- total of all salaries**

**FROM EMPLOYEE;**

**AVG(SALARY) MAX(SALARY) MIN(SALARY) SUM(SALARY)**

**----------- ----------- ----------- -----------**

**8775 24000 2500 175500**

**NEW PROBLEM:**

**How would we change the SQL for the President to show the same results but for Representatives only?**

Show the same for all Sales Reps and Marketing Reps.

SELECT AVG (salary), MAX (salary),

MIN (salary), SUM (salary)

FROM employee

WHERE job\_id LIKE '%REP%';

**AVG(SALARY) MAX(SALARY) MIN(SALARY) SUM(SALARY)**

**----------- ----------- ----------- -----------**

**8150 11000 6000 32600**

MORE COMPLEX PROBLEM: - already done this

Did this example before:

**The head of Payroll would like to know some information.**

**What is the minimum salary, average salary and the difference between the 2 values PLUS the highest salary and the difference between that and the average**

SELECT AVG(SALARY),

MIN(SALARY),

AVG(SALARY) - MIN(SALARY) "MIN. DIFF",

MAX(SALARY),

AVG(SALARY)- MAX(SALARY) "MAX DIFF"

FROM EMPLOYEE;

**AVG(SALARY) MIN(SALARY) MIN. DIFF MAX(SALARY) MAX DIFF**

**----------- ----------- ---------- ----------- ----------**

**8725 2500 6225 24000 -15275**

GUIDELINES

1

DISTINCT

- Makes the function consider only non-duplicate values

ALL

- Makes function consider every value

DEFAULT value is ALL and does not need to be specified

2

The DATA TYPES with the syntax ***expr*** argument may be

CHAR, VARCHAR2

NUMBER, DATE

3

All group functions ignore null values. To substitute for null values use NVL etc…

**Using MIN, MAX function examples**

You can use MIN and MAX for the following

- Numeric data

- Character data

- Date data

Find the newest and oldest employee by hire\_date

This shows the most senior employees, the one working the longest and the most junior employee

SELECT MIN(HIRE\_DATE),

MAX (HIRE\_DATE)

FROM EMPLOYEE;

MIN(HIRE\_ MAX(HIRE\_

--------- ---------

Poor titles due partly to column width too small to show title

17-JUN-87 29-JAN-00

Find the first person alphabetically by last name

Find the last employee by last name

🡺 function applied to character columns

SELECT **min** (last\_name) as "First in line",

**max** (last\_name) as "Always last to be called"

FROM employee;

First in line Always last to be called

------------------------- -------------------------

Abel Zlotkey

**Using the Count Function**

COUNT (\*) – returns the number of rows in a table

SELECT COUNT (\*)

FROM EMPLOYEE;

Using COUNT with an expression

SELECT COUNT (commission\_pct)

FROM EMPLOYEE

WHERE DEPARTMENT\_ID = 80;

COUNT(COMMISSION\_PCT)

---------------------

3

NOTE:

Count supplies the number of row that satisfies the SELECT statement in a table

Count with an expression returns the number of rows that meet the condition

select count(last\_name)

from employee

where last\_name between 'A' and 'G' 🡺 returns 5

Adding an expression returns non-null values

Adding DISTINCT returns the number of rows that are distinct from all the rows that are not null.

**Find the percentage of employees that receive a commission**

SELECT COUNT (\*),

COUNT(COMMISSION\_PCT), COUNT(COMMISSION\_PCT)/COUNT(\*)

FROM EMPLOYEE;

COUNT(\*) COUNT(COMMISSION\_PCT) COUNT(COMMISSION\_PCT)/COUNT(\*)

---------- --------------------- ------------------------------

21 5 0.2380952381

Needs improving

DISTINCT Examples

SELECT COUNT (DISTINCT department\_id)

FROM employee;

COUNT (DISTINCTDEPARTMENT\_ID)

----------------------------

7

Can you formulate the Question for this query?

**How many non-empty departments are in the company?**

**What is the average commission percent paid?**

GROUP FUNCTIONS and NULL

PROBLEM – average of just those who receive commission or of all employees

SELECT AVG (commission\_pct) -- just those getting commission

FROM employee;

AVG (COMMISSION\_PCT)

-------------------

.2125

SELECT AVG (NVL (commission\_pct, 0)) -- everybody

FROM employee;

AVG (NVL (COMMISSION\_PCT,0))

--------------------------

.0425

**Find the number of orders found in the order\_items table.**

**SELECT count (distinct order\_id)**

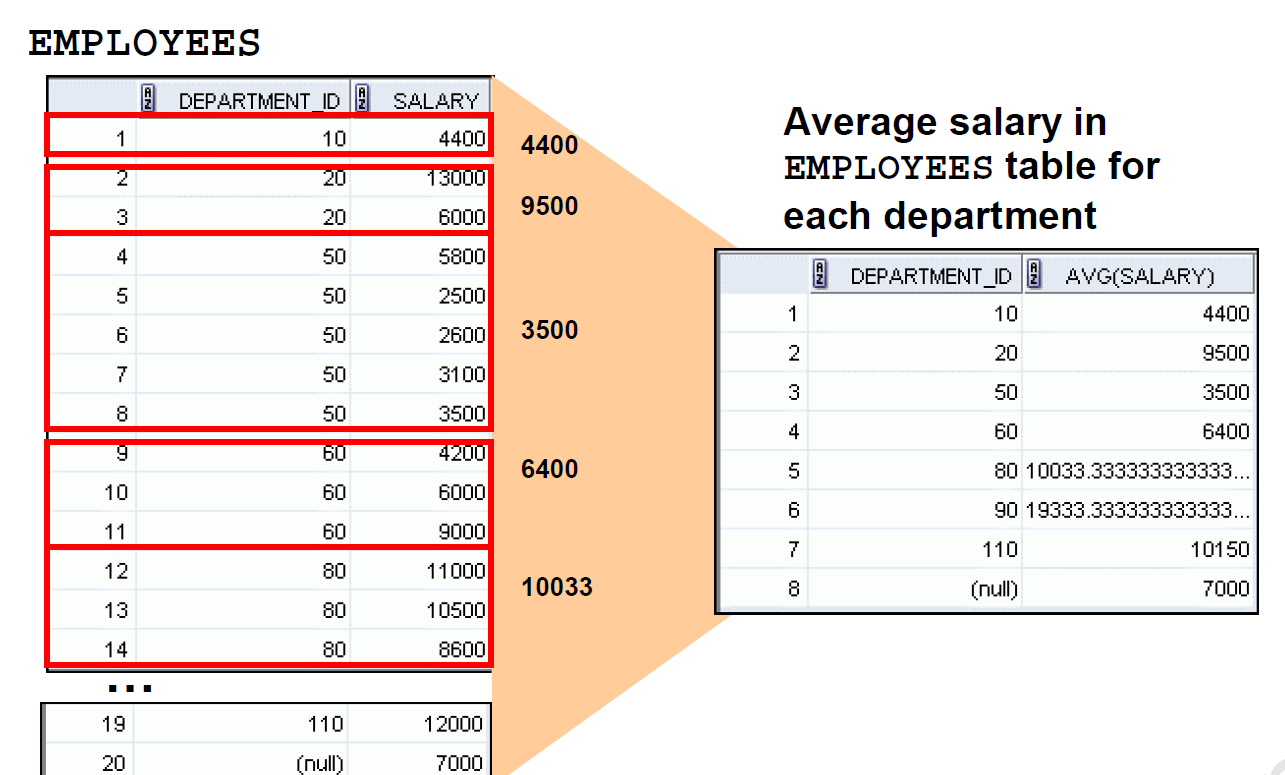
**FROM order\_items;**

Groups of Data

All group functions have treated the table so far as one large group

Sometimes the information needs to be divide into smaller groups

Example: Average by department



**GROUP BY**

Try this ----

SELECT DEPARTMENT\_ID, AVG (SALARY)

FROM EMPLOYEE;

ERROR at line 1:

ORA-00937: not a single-group group function

Why an error?

The use of department\_id results in a row of output for each row in the employee table

The AVG wants a single result for the entire table.

🡺 🡺 There is no sensible way to display that.

Introduces the GROUP BY to apply the group function by department\_id

SELECT DEPARTMENT\_ID, AVG (SALARY)

FROM EMPLOYEE

GROUP BY DEPARTMENT\_ID;

DEPARTMENT\_ID AVG(SALARY)

------------- -----------

7000

90 19333.3333

Rewrite the code to clean up the output

20 9500

110 10150

50 3500

80 10033.3333

60 6400

10 4400

8 rows selected

To clean up output

SELECT DEPARTMENT\_ID, round(AVG(SALARY),0)

FROM EMPLOYEE

GROUP BY DEPARTMENT\_ID;

BETTER

SELECT DEPARTMENT\_ID, round(AVG(SALARY),0)

FROM EMPLOYEE

GROUP BY DEPARTMENT\_ID

ORDER BY DEPARTMENT\_ID;

GROUP BY

The GROUP BY column does not need to be in the select

SELECT AVG(SALARY)

FROM EMPLOYEE

GROUP BY DEPARTMENT\_ID;

AVG(SALARY)

-----------

7000

19333.3333

Notice that the output is correct but is not very meaningful to the user without the department\_id showing

9500

10150

3500

10033.3333

6400

4400

GROUP BY often needs an ORDER BY

SELECT DEPARTMENT\_ID, AVG(SALARY)

FROM EMPLOYEE

GROUP BY DEPARTMENT\_ID

ORDER BY DEPARTMENT\_ID;

DEPARTMENT\_ID AVG(SALARY)

------------- -----------

10 4400

20 9500

Again, output should be cleaned up.

50 3500

60 6400

80 10033.3333

90 19333.3333

110 10150

7000

Grouping by more than 1 column

Groups within groups

PROBLEM:

**Display the total salary paid to each job title within each department**

LOGIC

Group employee by department

Within department group job titles

Sum up that lower level of grouping

SELECT department\_id, job\_id, SUM(salary)

FROM employee

GROUP BY department\_id, job\_id;

DEPARTMENT\_ID JOB\_ID SUM(SALARY)

------------- ---------- -----------

110 AC\_ACCOUNT 8300

90 AD\_VP 34000

Again, it is hard to see if it truly worked.

What would improve it?

50 ST\_CLERK 11700

80 SA\_REP 19600

50 ST\_MAN 5800

80 SA\_MAN 10500

110 AC\_MGR 12000

90 AD\_PRES 24000

60 IT\_PROG 19200

20 MK\_MAN 13000

SA\_REP 7000

10 AD\_ASST 4400

20 MK\_REP 6000

IMPROVED

**SELECT department\_id, job\_id, SUM(salary)**

**FROM employee**

**GROUP BY department\_id, job\_id**

**ORDER BY department\_id, job\_id**

DEPARTMENT\_ID JOB\_ID SUM(SALARY)

------------- ---------- -----------

10 AD\_ASST 4400

20 MK\_MAN 13000

20 MK\_REP 6000

50 ST\_CLERK 11700

50 ST\_MAN 5800

60 IT\_PROG 19200

80 SA\_MAN 10500

80 SA\_REP 19600

90 AD\_PRES 24000

90 AD\_VP 34000

110 AC\_ACCOUNT 8300

110 AC\_MGR 12000

SA\_REP 7000

13 rows selected

What would improve the report

DEPARTMENT\_ID JOB\_ID SUM(SALARY)

------------- ---------- -----------

10 AD\_ASST 4400

Improves readability for humans

20 MK\_MAN 13000

MK\_REP 6000

50 ST\_CLERK 11700

ST\_MAN 5800

60 IT\_PROG 19200

80 SA\_MAN 10500

SA\_REP 19600

90 AD\_PRES 24000

AD\_VP 34000

110 AC\_ACCOUNT 8300

AC\_MGR 12000

unk SA\_REP 7000

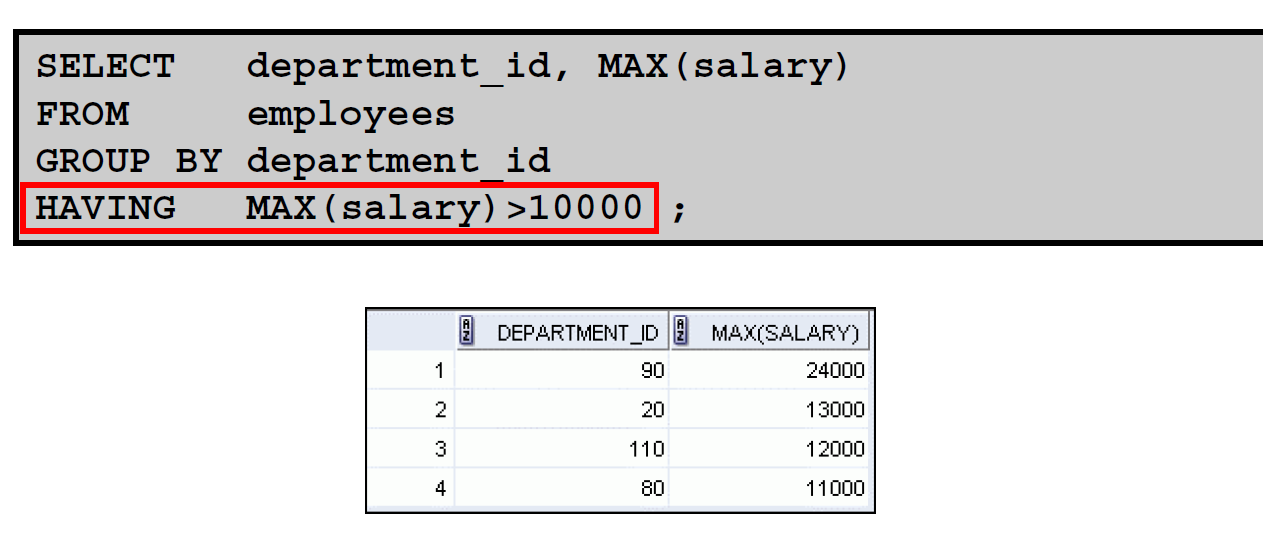
13 rows selected

Restricting Which Groups to Show

🡪**NOT** by using the WHERE clause

🡪 using the HAVING clause

**Find the maximum salary by department if maximum salary greater than 10,000**



Again nicer if put department in order

Add the ORDER BY clause

Nesting Group Functions

PROBLEM:

Display the department with the highest average salary

Logic:

Find the average salary by department

Find which one is the highest

SELECT MAX(AVG(salary))

FROM employee

GROUP BY department\_id;

MAX(AVG(SALARY))

----------------

19333.3333

PRACTICE: Last Page

**1 Write a query to determine how many job\_ids there are.**

**2 Write a query to find out how many people have the same job**

**3 Determine the number of managers (without listing them)**

**HINT: use the manager\_id**

**4 HR department want to know the range of salaries and what the difference is**

**1 Write a query to determine how many job\_ids there are.**

**SELECT count(distinct job\_id)**

**FROM employee;**

**2 Write a query to find out how many people have the same job**

**SELECT job\_id, count(\*)**

**FROM employee;**

**GROUP BY job\_id**

**3 Determine the number of managers (without listing them)**

**HINT: use the manager\_id**

**SELECT count(distinct manager\_id)**

**FROM employee;**

**4 HR department wants to know the range of salaries and what the difference is**

**SELECT max(salary),**

**min(salary),**

**max(salary)-min(salary) as "Difference"**

**FROM employee;**