

## SOBRE AGRUPAMIENTOS (GROUP BY)

Supongamos que nos interesa conocer cuantos empleados hay en el departamento 10

Dept_no	Emp_no	Apellido	Oficio	jefe	Fecha	Salario	comision
10	7782	CEREZO	DIRECTOR	7839	09/06/81	3185	
10	7839	REY	PRESIDENTE		17/11/81	6500	
10	7934	MUÑOZ	EMPLEADO	7782	23/01/82	1690	
20	7566	JIMENEZ	DIRECTOR	7839	02/04/81	3867	
20	7902	FERNANDEZ	ANALISTA	7566	03/12/81	3900	
20	7876	ALONSO	EMPLEADO	7788	23/09/81	1430	
20	7369	SANCHEZ	EMPLEADO	7788	17/12/80	1040	
20	7788	GIL	ANALISTA	7566	09/11/81	3900	
30	7521	SALA	VENDEDOR	7698	22/02/81	1625	650
30	7844	TOVAR	VENDEDOR	7698	08/09/81	1950	0
30	7499	ARROYO	VENDEDOR	7698	20/02/80	2080	390
30	7900	JIMENO	EMPLEADO	7698	03/12/81	1235	
30	7698	NEGRO	DIRECTOR	7839	01/05/81	3705	
30	7654	MARTIN	VENDEDOR	7698	29/09/81	1625	1820

**select \* from emple where dept\_no = 10** --> nos muestra los empleados del departamento 10

Dept_no	Emp_no	Apellido	Oficio	jefe	Fecha	Salario	comision
10	7782	CEREZO	DIRECTOR	7839	09/06/81	3185	
10	7839	REY	PRESIDENTE		17/11/81	6500	
10	7934	MUÑOZ	EMPLEADO	7782	23/01/82	1690	

si los contamos nos mostraría cuantos hay

**select count(emp\_no) from emple where dept\_no = 10**

Si queremos hacer lo mismo con los del departamento 20

**select \* from emple where dept\_no = 20** --> nos muestra los empleados del departamento 10

Dept_no	Emp_no	Apellido	Oficio	jefe	Fecha	Salario	comision
20	7566	JIMENEZ	DIRECTOR	7839	02/04/81	3867	
20	7902	FERNANDEZ	ANALISTA	7566	03/12/81	3900	
20	7876	ALONSO	EMPLEADO	7788	23/09/81	1430	
20	7369	SANCHEZ	EMPLEADO	7788	17/12/80	1040	
20	7788	GIL	ANALISTA	7566	09/11/81	3900	

si los contamos --> tenemos cuantos empleados hay en el 20

**select count(emp\_no) from emple where dept\_no = 20**

Esto es una forma de poder calcular cuántos empleados hay por cada departamento, pero poco operativa. Existe otra posibilidad que nos ofrece el SQL que es el agrupamiento con lo que podríamos hacer esto para todos los departamentos de una sola vez.

El group by lo que hace es crear subtablas *en memoria* en función del agrupamiento que ese indique, y sobre cada una de estas subtablas se evalúan las expresiones a visualizar

Dept_no	Emp_no	Apellido	Oficio	jefe	Fecha	Salario	comision
10	7782	CEREZO	DIRECTOR	7839	09/06/81	3185	
	7839	REY	PRESIDENTE		17/11/81	6500	
	7934	MUÑOZ	EMPLEADO	7782	23/01/82	1690	
20	7566	JIMENEZ	DIRECTOR	7839	02/04/81	3867	
	7902	FERNANDEZ	ANALISTA	7566	03/12/81	3900	
	7876	ALONSO	EMPLEADO	7788	23/09/81	1430	
	7369	SANCHEZ	EMPLEADO	7788	17/12/80	1040	
	7788	GIL	ANALISTA	7566	09/11/81	3900	
30	7521	SALA	VENDEDOR	7698	22/02/81	1625	650
	7844	TOVAR	VENDEDOR	7698	08/09/81	1950	0
	7499	ARROYO	VENDEDOR	7698	20/02/80	2080	390
	7900	JIMENO	EMPLEADO	7698	03/12/81	1235	
	7698	NEGRO	DIRECTOR	7839	01/05/81	3705	
	7654	MARTIN	VENDEDOR	7698	29/09/81	1625	1820

**IMPORTANTÍSIMO:** Cuando se hace un agrupamiento, se pueden visualizar en la sentencia select (como mucho):

- los campos que aparecen en el group by
- las expresiones que se evalúan de cada una de las subtablas
- nunca pueden aparecer campos/columnas a visualizar que no aparezcan en el group by (a menos que lo hagan en expresiones para la realización de cálculo p.e. count ( ... ) , sum(..), ... etc.

Si queremos contabilizar cuantos empleados hay en cada departamento

```
select dept_no, count(emp_no) from emple
       group by dept_no
```

DEPT_NO	COUNT(EMP_NO)
30	6
20	5
10	3

aunque incluimos un campo -emp\_no- que no está en el agrupamiento, esté se encuentra en una expresión, nunca podríamos usar el emp\_no solo sin expresión ya que no se encuentra en el group by. En esta sentencia indicamos que para cada departamento cuente cuantos empleados tiene y muestre el número de departamento y el número de empleados que tiene

Sin embargo, en la siguiente sentencia:

```
select dept_no, emp_no from emple
       group by dept_no
```

Esta sentencia produciría un error ya que el campo emp\_no no forma parte del group by. Si analizamos bien la expresión estamos diciendo que agrupemos las filas de la tabla por dept\_no, pero estamos indicándole que muestre para cada departamento (dept\_no) y empleado (emp\_no) y agrupándolo por departamento → esto no tiene sentido, no hay concordancia.

Recordar: como mucho, solo podremos visualizar de todas las columnas aquellas que forman parte del agrupamiento y expresiones a calcular sobre cada una de las subtablas generadas por el agrupamiento

Por ejemplo si queremos saber cuantos empleados y cual es la media de los salarios en cada departamento

```
select dept_no, count(emp_no), avg(salario) from emple
       group by dept_no
```

DEPT_NO	COUNT(EMP_NO)	TRUNC(AVG(SALARIO),2)
30	6	2036.66
20	5	2827.4
10	3	3791.66

Supongamos que queremos saber cuantos empleados de cada oficio hay por departamento, tendríamos que partir de una tabla en memoria agrupada por departamento y oficio

Dept_no	Oficio	Emp_no	Apellido	jefe	Fecha	Salario	comision
10	DIRECTOR	7782	CEREZO	7839	09/06/81	3185	
10	PRESIDENTE	7839	REY		17/11/81	6500	
10	EMPLEADO	7934	MUÑOZ	7782	23/01/82	1690	
20	DIRECTOR	7566	JIMENEZ	7839	02/04/81	3867	
20	ANALISTA	7788	GIL	7566	09/11/81	3900	
	ANALISTA	7902	FERNANDEZ	7566	03/12/81	3900	
20	EMPLEADO	7876	ALONSO	7788	23/09/81	1430	
	EMPLEADO	7369	SANCHEZ	7788	17/12/80	1040	
30	VENDEDOR	7521	SALA	7698	22/02/81	1625	650
	VENDEDOR	7844	TOVAR	7698	08/09/81	1950	0
	VENDEDOR	7499	ARROYO	7698	20/02/80	2080	390
	VENDEDOR	7654	MARTIN	7698	29/09/81	1625	1820
30	EMPLEADO	7900	JIMENO	7698	03/12/81	1235	
30	DIRECTOR	7698	NEGRO	7839	01/05/81	3705	

```
select dept_no, oficio, count(emp_no) from emple
       group by dept_no, oficio
```

DEPT_NO	OFICIO	COUNT(EMP_NO)
10	DIRECTOR	1
10	EMPLEADO	1
10	PRESIDENTE	1
20	ANALISTA	2
20	DIRECTOR	1
20	EMPLEADO	2
30	DIRECTOR	1
30	EMPLEADO	1
30	VENDEDOR	4

Como vemos tenemos que agrupar por departamento y por oficio y como mucho estos serán las columnas que podremos visualizar de todas las columnas de la tabla empleados, lo que sí podremos hacer es visualizar evaluar expresiones sobre el resto de las columnas de cada una de las subtablas dentro del agrupamiento; por ejemplo `count(emp_no)` o por ejemplo si quisiéramos saber para cada departamento y dentro de este para cada oficio cual es el máximo salario que hay podríamos calcular el `max(salario)` de cada una de las subtablas.

## CLAUSULA HAVING

**HAVING:** --> Su función es parecida al `WHERE`, pero sobre el resultado del agrupamiento, es decir, una vez realizada la consulta con el agrupamiento, si existe un `having` se selecciona del resultado del agrupamiento aquellas filas que cumplan la condición indicada en el `having`.

Si queremos contabilizar cuantos empleados hay en cada departamento

```
select dept_no, count(emp_no) from emple
       group by dept_no
```

DEPT_NO	COUNT(EMP_NO)
30	6
20	5
10	3

Si además queremos saber solo aquellos departamentos que tienen más de 5 empleados , entonces

```
select dept_no, count(emp_no) from emple
       group by dept_no having count(emp_no) >= 5
```

DEPT_NO	COUNT(EMP_NO)
30	6
20	5

Es decir, el `HAVING` solo actuará sobre la tabla resultante del agrupamiento

## ORDEN DE EVALUACIÓN DE LAS CLAUSULAS DEL SELECT

Un select con varias cláusulas se evaluará según el siguiente orden:

**SELECT..... FROM....**

**WHERE...**

**[GROUP BY colum1, colum2,...] [HAVING condicion]**

**[ORDER BY.....]**

- |    |   |          |   |
|----|---|----------|---|
| 1º | ↓ | WHERE    | : Selecciona las filas                    |
| 2º |   | GROUP BY | : Agrupa estas filas                      |
| 3º |   | HAVING   | : Filtra los grupos. Selecciona y elimina |
| 4º | ↓ | ORDER BY | : Clasifica la salida. Ordena los grupos  |

## SUBCONSULTAS CON SENTENCIAS SELECT CON GROUP BY

Una sentencia select con agrupamiento puede ser utilizada como subconsulta de otra consulta.

Ejemplo: Cual es el departamento que mas empleados tiene

Si queremos contabilizar cuantos empleados hay en cada departamento

```
select dept_no, count(emp_no) from emple
      group by dept_no
```

DEPT_NO	COUNT(EMP_NO)
30	6
20	5
10	3

Yo puedo calcular el máximo de la columna count(emp\_no)

```
select max(count(emp_no)) from emple
      group by dept_no
```

COUNT(EMP_NO)
6

A partir de aquí yo podría utilizar la primera consulta para conocer cual es dept\_no cuyo count(dept\_no) coincide con el valor calculado anteriormente, es decir cual es la fila cuyo count(emp\_no) es igual a ese máximo calculado antes

```
select dept_no, count(emp_no) from emple
      group by dept_no
      having count(emp_no) = (select max(count(emp_no)) from emple
                             group by dept_no)
```

DEPT_NO	COUNT(EMP_NO)
30	6