

Table of Contents

1. The Cartesian Product or Cross Join..... 1

With an inner Join between tables we define how the tables are to be related to each other. With a Cartesian product we Join two tables but without any attempt to relate the tables to each other. Most of the time a Cartesian product is not what you want to use, but over the semester we will find places where the Cartesian product is helpful.

1. The Cartesian Product or Cross Join

Assume we have the following tables and rows. The SQL is in the demo

Z_EM_Dept		Z_EM_Emp			Z_EM_EmpProj	
D_ID	D_Name	E_ID	E_Name	D_ID	P_ID	E_ID
100	Manufacturing	1	Jones	150	ORDB-10	3
150	Accounting	2	Martin	150	ORDB-10	5
200	Marketing	3	Gates	250	Q4-SALES	2
250	Research	4	Anders	100	Q4-SALES	4
		5	Bossy		ORDB-10	2
		6	Perkins		Q4-SALES	5

Suppose I want to display the name of every employee who is assigned to a department and the name of that department. You should be thing that this is done with an inner join- and that would be correct.

Demo 01: Inner join.

```
Select
  D.d_name, D.d_id
, E.d_id, E.e_id, E.e_name
From z_em_dept D
Join z_em_emp E on D.d_id = E.d_id;
```

d_name	d_id	d_id	e_id	e_name
Accounting	150	150	1	Jones
Accounting	150	150	2	Martin
Research	250	250	3	Gates
Manufacturing	100	100	4	Anders

The result set contain the department id twice- once From the department table and once From the employee table.

Employee id 1 works in dept 150 which is the Accounting department.

But suppose you were in a hurry and you made an error. In the From clause you listed the two tables with a comma between them and skipped the On clause.

Demo 02: Getting a Cartesian product by default.

```

Select
  z_em_dept.d_name, z_em_dept.d_id
, z_em_emp.d_id, z_em_emp.e_id, z_em_emp.e_name
From z_em_dept
, z_em_emp;

```

d_name	d_id	d_id	e_id	e_name
Manufacturing	100	150	1	Jones
Accounting	150	150	1	Jones
Marketing	200	150	1	Jones
Research	250	150	1	Jones
Manufacturing	100	150	2	Martin
Accounting	150	150	2	Martin
Marketing	200	150	2	Martin
Research	250	150	2	Martin
Manufacturing	100	250	3	Gates
Accounting	150	250	3	Gates
Marketing	200	250	3	Gates
Research	250	250	3	Gates
Manufacturing	100	150	4	Anders
Accounting	150	150	4	Anders
Marketing	200	150	4	Anders
Research	250	150	4	Anders
Manufacturing	100	NULL	5	Bossy
Accounting	150	NULL	5	Bossy
Marketing	200	NULL	5	Bossy
Research	250	NULL	5	Bossy
Manufacturing	100	NULL	6	Perkins
Accounting	150	NULL	6	Perkins
Marketing	200	NULL	6	Perkins
Research	250	NULL	6	Perkins

24 rows in set (0.00 sec)

You get 24 rows in the result set and it looks like employee 1 works in Manufacturing and in Accounting and in Marketing and in Research. But it looks like employee 2 also works in all of those departments- In fact every employee seems to work in very department including Bossy who is assigned to no department at all.

This is a legal query- it runs without error and produces a result set; but the result set is not very meaningful- unless you wanted to see a table of all possible assignments of employees to all departments.

What you have here is a Cartesian product of two tables; it contains each row in the first table associated with each of the rows in the second table. There is no Join on matching attribute values. Therefore, we get each of the 4 department rows matched with each of the 6 employee rows for a total of 24 rows.

Demo 03: Specifying the cross Join gives the same results but makes it obvious that we are intentionally doing a Cartesian product.

```

Select z_em_dept.d_name, z_em_dept.d_id
, z_em_emp.d_id, z_em_emp.e_id, z_em_emp.e_name
From z_em_dept
CROSS JOIN z_em_emp;

```

Demo 04: If we added in the project table with no joins, then we get 144 rows returned.

```
Select z_em_dept.d_name, z_em_dept.d_id
, z_em_emp.d_id, z_em_emp.e_id, z_em_emp.e_name
, p_id
From z_em_dept
CROSS JOIN z_em_emp
CROSS JOIN z_em_EmpProj;
```

d_name	d_id	d_id	e_id	e_name	p_id
Manufacturing	100	150	1	Jones	ORDB-10
Accounting	150	150	1	Jones	ORDB-10
Marketing	200	150	1	Jones	ORDB-10
Research	250	150	1	Jones	ORDB-10
Manufacturing	100	150	2	Martin	ORDB-10
Accounting	150	150	2	Martin	ORDB-10
Marketing	200	150	2	Martin	ORDB-10
Research	250	150	2	Martin	ORDB-10
Manufacturing	100	250	3	Gates	ORDB-10
Accounting	150	250	3	Gates	ORDB-10
Marketing	200	250	3	Gates	ORDB-10
Research	250	250	3	Gates	ORDB-10
Manufacturing	100	150	4	Anders	ORDB-10
Accounting	150	150	4	Anders	ORDB-10
Marketing	200	150	4	Anders	ORDB-10
Research	250	150	4	Anders	ORDB-10
Manufacturing	100	NULL	5	Bossy	ORDB-10
Accounting	150	NULL	5	Bossy	ORDB-10
Marketing	200	NULL	5	Bossy	ORDB-10
Research	250	NULL	5	Bossy	ORDB-10

. . . rows omitted

144 rows would take a bit more than two pages to display- that is a lot of rows and these are tiny tables.

Suppose you did a cross Join of the vets clients table (14 rows) with the animal table (29 rows) The result set would have 406 rows. Now consider adding the exam headers table (26 rows) to the From clause The result set would now have 10,556 rows. Of those rows only 26 are meaningful (in the sense that they matching up a client to an animal to an exam). If you think of 60 lines to a page, 10,556 rows would take 176 pages to print.

The number of rows in the result set of a Cartesian product is the product of the number of rows in each table.

Suppose our clinic had 300 clients with an average of 2 animals each and that there were an average of 2 exams per animal each year.

Our tables would have 300 rows for clients, 600 rows for animals and 1200 rows for exams. If we wanted to display exam header data for last year we would expect about 1200 rows in the result table. With a Cartesian product we would get 216,000,000 rows in the result set

$300 * 600 * 1200 = 216000000$

Sometimes there is a value in intentionally doing a Cartesian product of two tables, but most of the time a query with a Cartesian product has been miscoded. If your query result sets has a lot more rows than you expected, check your joins carefully.