Inner Joins Gone Wrong

SQL

A Brief Explanation About How SQL Performs Inner Joins and How to Not Mess It Up

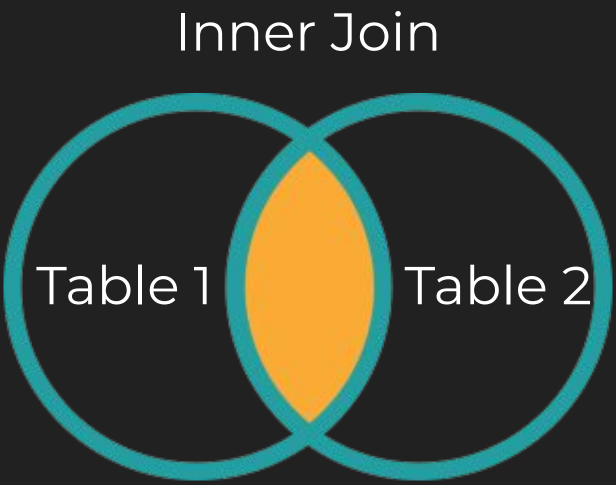
By Jeffrey Wigington

Introduction

An Inner Join returns only rows where the selected column have matching values.

But that doesn't mean you always end up with a smaller table after the join!

That's the concept I want to teach you now.



Brief Mathematical Overview

DON'T LEAVE YET!

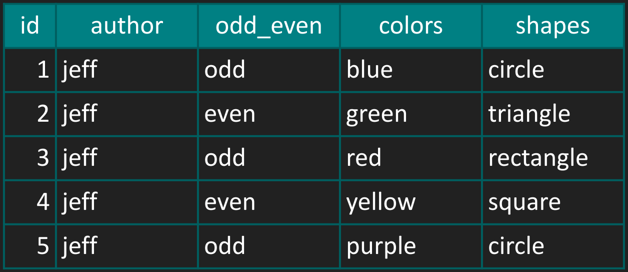
You just need to understand that when SQL looks for a value in Table 2 from Table 1, it doesn't just grab the first record it finds. It finds ALL of the records that match that value.

Every. Time.

If Table 1 has M rows and Table 2 has N rows, the resulting Joined table can have 0 rows (no matching values) to M*N rows (only matching values). This can make for some ***insanely*** large tables as a result.

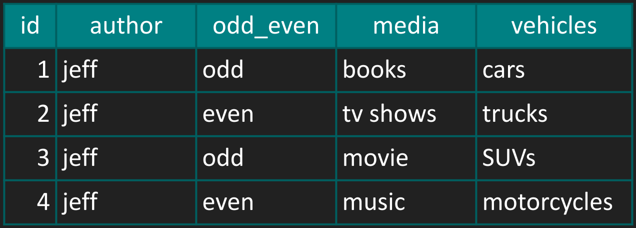
Phew. Time to Look at Data!

Table 1



id	author	odd_even	colors	shapes
1	jeff	odd	blue	circle
2	jeff	even	green	triangle
3	jeff	odd	red	rectangle
4	jeff	even	yellow	square
5	jeff	odd	purple	circle

Table 2



id	author	odd_even	media	vehicles
1	jeff	odd	books	cars
2	jeff	even	tv shows	trucks
3	jeff	odd	movie	SUVs
4	jeff	even	music	motorcycles

Table 1 is 5 rows x 5 columns
Table 2 is 4 rows x 5 columns

We expect between 0 and 5*4 (20) rows.

The next couple pages will show the resulting Inner Joins over the first three columns:

- id (primary key - therefore unique values)
- author (foreign key - all the same value)
- odd_even (foreign key - half and half same values)

This will illustrate what happens if you try to Inner Join on the wrong attribute of your data. Hint: you don’t want to mess this up

Inner Join on “id”

Table 1

id	author	odd_even	colors	shapes
1	jeff	odd	blue	circle
2	jeff	even	green	triangle
3	jeff	odd	red	rectangle
4	jeff	even	yellow	square
5	jeff	odd	purple	circle

Table 2

id	author	odd_even	media	vehicles
1	jeff	odd	books	cars
2	jeff	even	tv shows	trucks
3	jeff	odd	movie	SUVs
4	jeff	even	music	motorcycles

Inner Join on "id"

id	author	odd_even	colors	shapes	id	author	odd_even	media	vehicles
1	jeff	odd	blue	circle	1	jeff	odd	books	cars
2	jeff	even	green	triangle	2	jeff	even	tv shows	trucks
3	jeff	odd	red	rectangle	3	jeff	odd	movie	SUVs
4	jeff	even	yellow	square	4	jeff	even	music	motorcycles

Since each value is unique, it’s a one-for-one join until Table 2 runs out of rows. If Table 2 had an “id” of value “7”, it would be excluded.

Inner Join on “author”

Table 1

id	author	odd_even	colors	shapes
1	jeff	odd	blue	circle
2	jeff	even	green	triangle
3	jeff	odd	red	rectangle
4	jeff	even	yellow	square
5	jeff	odd	purple	circle

Table 2

id	author	odd_even	media	vehicles
1	jeff	odd	books	cars
2	jeff	even	tv shows	trucks
3	jeff	odd	movie	SUVs
4	jeff	even	music	motorcycles

Since each value in “author” is the same, for every row of Table 1, all 4 rows of Table 2 are joined with it for every time.

This results in a joined table of size 20 rows (5*4) by 10 columns.

id	author	odd_even	colors	shapes	id	author	odd_even	media	vehicles
1	jeff	odd	blue	circle	1	jeff	odd	books	cars
1	jeff	odd	blue	circle	2	jeff	even	tv shows	trucks
1	jeff	odd	blue	circle	3	jeff	odd	movie	SUVs
1	jeff	odd	blue	circle	4	jeff	even	music	motorcycles
2	jeff	even	green	triangle	1	jeff	odd	books	cars
2	jeff	even	green	triangle	2	jeff	even	tv shows	trucks
2	jeff	even	green	triangle	3	jeff	odd	movie	SUVs
2	jeff	even	green	triangle	4	jeff	even	music	motorcycles
3	jeff	odd	red	rectangle	1	jeff	odd	books	cars
3	jeff	odd	red	rectangle	2	jeff	even	tv shows	trucks
3	jeff	odd	red	rectangle	3	jeff	odd	movie	SUVs
3	jeff	odd	red	rectangle	4	jeff	even	music	motorcycles
4	jeff	even	yellow	square	1	jeff	odd	books	cars
4	jeff	even	yellow	square	2	jeff	even	tv shows	trucks
4	jeff	even	yellow	square	3	jeff	odd	movie	SUVs
4	jeff	even	yellow	square	4	jeff	even	music	motorcycles
5	jeff	odd	purple	circle	1	jeff	odd	books	cars
5	jeff	odd	purple	circle	2	jeff	even	tv shows	trucks
5	jeff	odd	purple	circle	3	jeff	odd	movie	SUVs
5	jeff	odd	purple	circle	4	jeff	even	music	motorcycles

Inner Join on “odd_even”

Table 1

id	author	odd_even	colors	shapes
1	jeff	odd	blue	circle
2	jeff	even	green	triangle
3	jeff	odd	red	rectangle
4	jeff	even	yellow	square
5	jeff	odd	purple	circle

Table 2

id	author	odd_even	media	vehicles
1	jeff	odd	books	cars
2	jeff	even	tv shows	trucks
3	jeff	odd	movie	SUVs
4	jeff	even	music	motorcycles

Inner Join on "odd_even"

id	author	odd_even	colors	shapes	id	author	odd_even	media	vehicles
1	jeff	odd	blue	circle	1	jeff	odd	books	cars
1	jeff	odd	blue	circle	3	jeff	odd	movie	SUVs
2	jeff	even	green	triangle	2	jeff	even	tv shows	trucks
2	jeff	even	green	triangle	4	jeff	even	music	motorcycles
3	jeff	odd	red	rectangle	1	jeff	odd	books	cars
3	jeff	odd	red	rectangle	3	jeff	odd	movie	SUVs
4	jeff	even	yellow	square	2	jeff	even	tv shows	trucks
4	jeff	even	yellow	square	4	jeff	even	music	motorcycles
5	jeff	odd	purple	circle	1	jeff	odd	books	cars
5	jeff	odd	purple	circle	3	jeff	odd	movie	SUVs

Table 1 has three “odd” and two “even”. Table 2 has two of each “odd” and “even”.

That’s 3x2+2x2, which results in a joined table of 10 rows by 10 columns.

Inner Join on Other Columns?

There's no point in illustrating the Inner Join on the other columns because they would result in an empty joined table. For completeness however, I've included the joined table below. As you can see, it's empty.

The Takeaway?

There are two major points I want you to take away from this:

1. Stick with Primary Keys and/or columns with unique values when doing Joins (it will affect all other types of Joins too).

2. Take time to anticipate your results before you run your code to verify that the results make sense and that it’s what you would expect.

Congratulations! You Made It!

Since you made it this far, you’re either bored of scrolling through LinkedIn or you’re pretty interested in Data Analytics.

Maybe you want to see more of my stuff? Well, here’s a link to my portfolio site where you can see some of my work. Before you leave, drop a “like” or a comment below if you found this helpful.

Thanks for the support!
- Jeffrey Wigington