

### Spark Joins that have a SQL Equivalent

#### Basic Syntax:

```
df1.join(df2, df1("column_name_from_m1") === df2("column_name_from_m2"), "method").show()
//.show(false)
```

#### Self:

Though there is no self-join type available, we can use any of the above-explained join types to join DataFrame to itself.

```
println("self join")
empDF.as("emp1").join(empDF.as("emp2"),
  col("emp1.sup_emp_id") === col("emp2.emp_id"), "inner")
.select(col("emp1.emp_id"), col("emp1.name"),
  col("emp2.emp_id").as("sup_emp_id"),
  col("emp2.name").as("sup_emp_name"))
.show(false)
```

#### Inner:

```
empDF.join(deptDF, empDF("emp_dept_id") === deptDF("dept_id"), "inner")
.show(false)
```

#### Full outer:

```
empDF.join(deptDF, empDF("emp_dept_id") === deptDF("dept_id"), "full")
.show(false)
```

#### Left Outer:

```
empDF.join(deptDF, empDF("emp_dept_id") === deptDF("dept_id"), "left")
.show(false)
```

### Spark Joins that are not in SQL!

#### Left semi join

Returns all rows from the left DF on records match in the right dataset on join expression, records not matched on join expression are ignored from both left and right datasets.

```
empDF.join(deptDF, empDF("emp_dept_id") === deptDF("dept_id"), "leftsemi")
.show(false)
```

#### Left anti join

leftanti join returns only columns from the left DataFrame / Dataset for non-matched records.

```
empDF.join(deptDF, empDF("emp_dept_id") === deptDF("dept_id"), "leftanti")
.show(false)
```

#### Cross Join

use the CROSS JOIN syntax to allow cartesian products between these relations, or: enable implicit cartesian products by setting the configuration variable `spark.sql.crossJoin.enabled=true`;

```
println("cross join")
empDF.join(deptDF, empDF("emp_dept_id") === deptDF("dept_id"), "cross")
.show(false)
println("Using crossJoin()")
```



Spark Joins that are not in SQL! (cont)

```
> empDF.crossJoin(deptDF).show(false)
```



By **datamansam**

[cheatography.com/datamansam/](https://cheatography.com/datamansam/)

Not published yet.

Last updated 29th September, 2022.

Page 3 of 3.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>

