# Spark SQL Joins Cheat Sheet

## Spark Joins that have a SQL Equivalent

```
Basic Syntax:
df1.jo in(df2, df1("co mmo n_c ol_ fro m1") === df2("co mmo n_c ol_ fro m2"),
" met hod " ).s how ()/.sh -
ow( false)
Self:
   Though there is no self-join type available, we can use any of the above- exp lained join types
to join DataFrame to itself.
   pri ntl n("self join")
     emp DF.a s( " emp 1").j oi n(e mpD F.a s("e mp2 "),
         col ("em p1.s up _em p_i d") === col("em p2.e mp _id " ),"i nne r")
        .se lec t(c ol( " emp 1.e mp_ id") ,co l("e mp1.na me"),
             col ("em p2.e mp _id " ).a s("s up_ emp _id "),
             col ("em p2.n am e").a s( " sup _em p_n ame "))
            .sh ow( false)
Inner:
     emp DF.j oi n(d ept DF, emp DF( " emp _de pt_ id") === deptDF ("de pt_ id") ,"in ner ")
        .sh ow( false)
Full outer:
         emp DF.j oi n(d ept DF, emp DF( " emp _de pt_ id") ===
   deptDF ("de pt_ id") ,"fu ll")
        .sh ow( false)
Left Outer:
empDF.j oi n(d ept DF, emp DF( " emp _de pt_ id") === deptDF ("de pt_ id") ,"le ft")
        .sh ow( 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 expres sion, records
not matched on join expression are ignored from both left and right datasets.

empDF.j oi n(d ept DF, emp DF( " emp _de pt_ id") === deptDF ("de pt_ id") ,"le fts emi ")


   .sh ow( false)
Left anti join
leftanti join returns only columns from the left DataFr ame /Da taset for non-ma tched
records. empDF.j oi n(d ept DF, emp DF( " emp _de pt_ id") ===
deptDF ("de pt_ id") ,"le fta nti ")
        .sh ow( false)

use the CROSS JOIN syntax to allow cartesian products between these
relations, or: enable implicit cartesian products by setting the
config uration variable spark.s ql.cr oss Joi n.e nab led =true;
   pri ntl n("cross join")
     emp DF.j oi n(d ept DF, emp DF( " emp _de pt_ id") === deptDF ("de pt_ id") ,"cr oss ")
        .sh ow( false)
     pri ntl n("Using crossJ oin ()")
```

# Spark SQL Joins Cheat Sheet

## Spark Joins that are not in SQL! (cont)

```
emp  DF.c  ro  ssJ  oin  (de  ptD  F).s  ho  w(f  alse)
```