```
# Databricks notebook source
import pyspark
from pyspark.sql import SparkSession
spark =
SparkSession.builder.appName('machinelearninggeeks.com').getOrCreate()

emp = [(1,"Smith",-1,"2018","10","M",3000), \
    (2,"Rose",1,"2010","20","M",4000), \
    (3,"Williams",1,"2010","10","M",1000), \
    (4,"Jones",2,"2005","10","F",2000), \
    (5,"Brown",2,"2010","40","",-1), \
      (6,"Brown",2,"2010","50","",-1) \
  ]
empColumns = ["emp_id","name","superior_emp_id","year_joined", \
       "emp_dept_id","gender","salary"]

empDF = spark.createDataFrame(data=emp, schema = empColumns)
empDF.printSchema()
empDF.show(truncate=False)

dept = [("Finance",10), \
    ("Marketing",20), \
    ("Sales",30), \
    ("IT",40) \
  ]
deptColumns = ["dept_name","dept_id"]
deptDF = spark.createDataFrame(data=dept, schema = deptColumns)
deptDF.printSchema()
deptDF.show(truncate=False)

empDF.join(deptDF,empDF.emp_dept_id ==
deptDF.dept_id,"left").show(truncate=False)
empDF.join(deptDF,empDF.emp_dept_id ==
deptDF.dept_id,"leftouter").show(truncate=False)

empDF.createOrReplaceTempView("EMP")
deptDF.createOrReplaceTempView("DEPT")

joinDF2 = spark.sql("SELECT e.* FROM EMP e LEFT ANTI JOIN DEPT d ON
e.emp_dept_id == d.dept_id") \
  .show(truncate=False)

# COMMAND ----------

i
```