```python
# Databricks notebook source
from pyspark.sql import SparkSession

# Create SparkSession
spark = SparkSession.builder \
         .appName('SparkByExamples.com') \
         .getOrCreate()
#EMP DataFrame
empData = [(1,"Smith",10), (2,"Rose",20),
    (3,"Williams",10), (4,"Jones",30)
  ]
empColumns = ["emp_id","name","emp_dept_id"]
empDF = spark.createDataFrame(empData,empColumns)
empDF.show()

#DEPT DataFrame
deptData = [("Finance",10), ("Marketing",20),
    ("Sales",30),("IT",40)
  ]
deptColumns = ["dept_name","dept_id"]
deptDF=spark.createDataFrame(deptData,deptColumns)
deptDF.show()

#Address DataFrame
addData=[(1,"1523 Main St","SFO","CA"),
    (2,"3453 Orange St","SFO","NY"),
    (3,"34 Warner St","Jersey","NJ"),
    (4,"221 Cavalier St","Newark","DE"),
    (5,"789 Walnut St","Sandiago","CA")
  ]
addColumns = ["emp_id","addline1","city","state"]
addDF = spark.createDataFrame(addData,addColumns)
addDF.show()

#Join two DataFrames
empDF.join(addDF,empDF["emp_id"] == addDF["emp_id"]).show()

#Drop duplicate column
empDF.join(addDF,["emp_id"]).show()

#Join Multiple DataFrames
empDF.join(addDF,["emp_id"]) \
    .join(deptDF,empDF["emp_dept_id"] == deptDF["dept_id"]) \
    .show()

#Using Where for Join Condition
empDF.join(deptDF).where(empDF["emp_dept_id"] == deptDF["dept_id"]) \
    .join(addDF).where(empDF["emp_id"] == addDF["emp_id"]) \
    .show()
```

```
#SQL
empDF.createOrReplaceTempView("EMP")
deptDF.createOrReplaceTempView("DEPT")
addDF.createOrReplaceTempView("ADD")

spark.sql("select * from EMP e, DEPT d, ADD a " + \
    "where e.emp_dept_id == d.dept_id and e.emp_id == a.emp_id") \
    .show()

#
df1 = spark.createDataFrame(
    [(1, "A"), (2, "B"), (3, "C")],
    ["A1", "A2"])

df2 = spark.createDataFrame(
    [(1, "F"), (2, "B")],
    ["B1", "B2"])

df = df1.join(df2, (df1.A1 == df2.B1) & (df1.A2 == df2.B2))
df.show()

# COMMAND ----------
```