# Session 21 Assignment 2

Import the flight details list

val flights = sc.parallelize(List(

| ("SEA", "JFK", "DL", "418", "7:00"),

| ("SFO", "LAX", "AA", "1250", "7:05"),

| ("SFO", "JFK", "VX", "12", "7:05"),

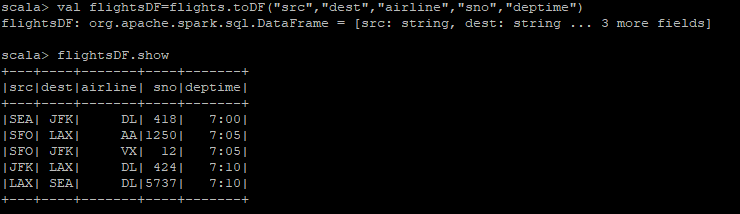
| ("JFK", "LAX", "DL", "424", "7:10"),

| ("LAX", "SEA", "DL", "5737", "7:10")))

Convert the flight details for DataFrame

val flightsDF=flights.toDF("src","dest","airline","sno","duration")

flightsDF.show



Convert the airport details list to DataFrame

val airports = sc.parallelize(List(

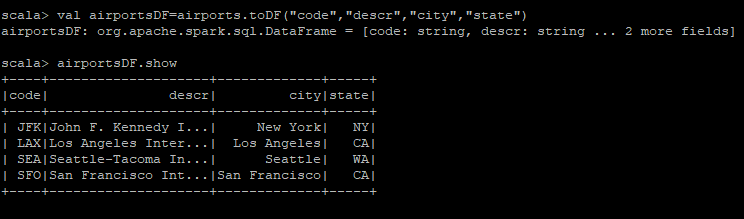
("JFK", "John F. Kennedy International Airport", "New York", "NY"),

("LAX", "Los Angeles International Airport", "Los Angeles", "CA"),

("SEA", "Seattle-Tacoma International Airport", "Seattle", "WA"),

("SFO", "San Francisco International Airport", "San Francisco", "CA")))

val airportsDF=airports.toDF("code","descr","city","state")



Convert the airlines details to DataFrame

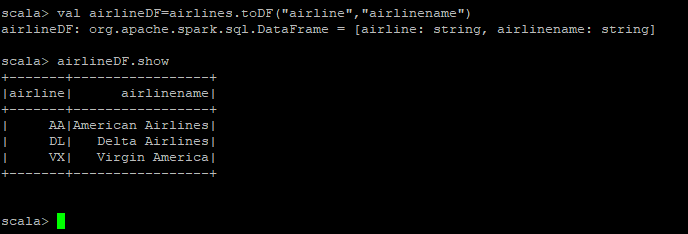
val airlines = sc.parallelize(List(

("AA", "American Airlines"),

("DL", "Delta Airlines"),

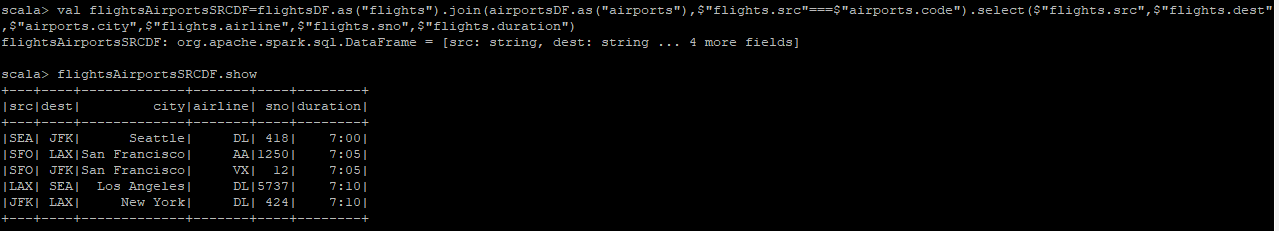
("VX", "Virgin America")))

val airlineDF=airlines.toDF("airline","airlinename")



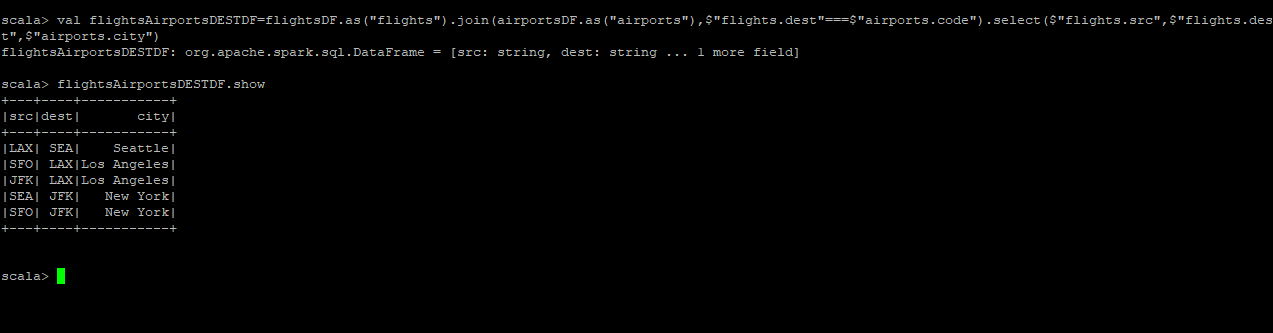
Step 1: Join flights dataframe and airport dataframe to get the source city.

val flightsAirportsSRCDF=flightsDF.as("flights").join(airportsDF.as("airports"),$"flights.src"===$"airports.code").select($"flights.src",$"flights.dest",$"airports.city",$"flights.airline",$"flights.sno",$"flights.duration")



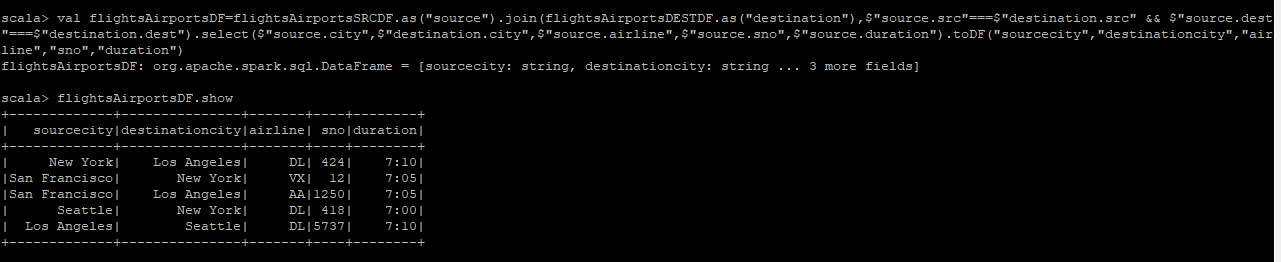
Step 2: Again, perform the same join as above to get the destination city.

val flightsAirportsDESTDF=flightsDF.as("flights").join(airportsDF.as("airports"),$"flights.dest"===$"airports.code").select($"flights.src",$"flights.dest",$"airports.city")



Step 3: Again, perform join on dataframes in step 1 and 2 to get the source city and destination city.the same join as above to get the destination city.

val flightsAirportsDESTDF=flightsDF.as("flights").join(airportsDF.as("airports"),$"flights.dest"===$"airports.code").select($"flights.src",$"flights.dest",$"airports.city")



Step 4: Now, to get the final result make a join of step 3 with airline dataframe

val result=flightsAirportsDF.alias("f").join(airlineDF.alias("a"),$"a.airline"===$"f.airline").select($"f.sourcecity",$"f.destinationcity",$"a.airlinename",$"f.sno",$"f.duration")

