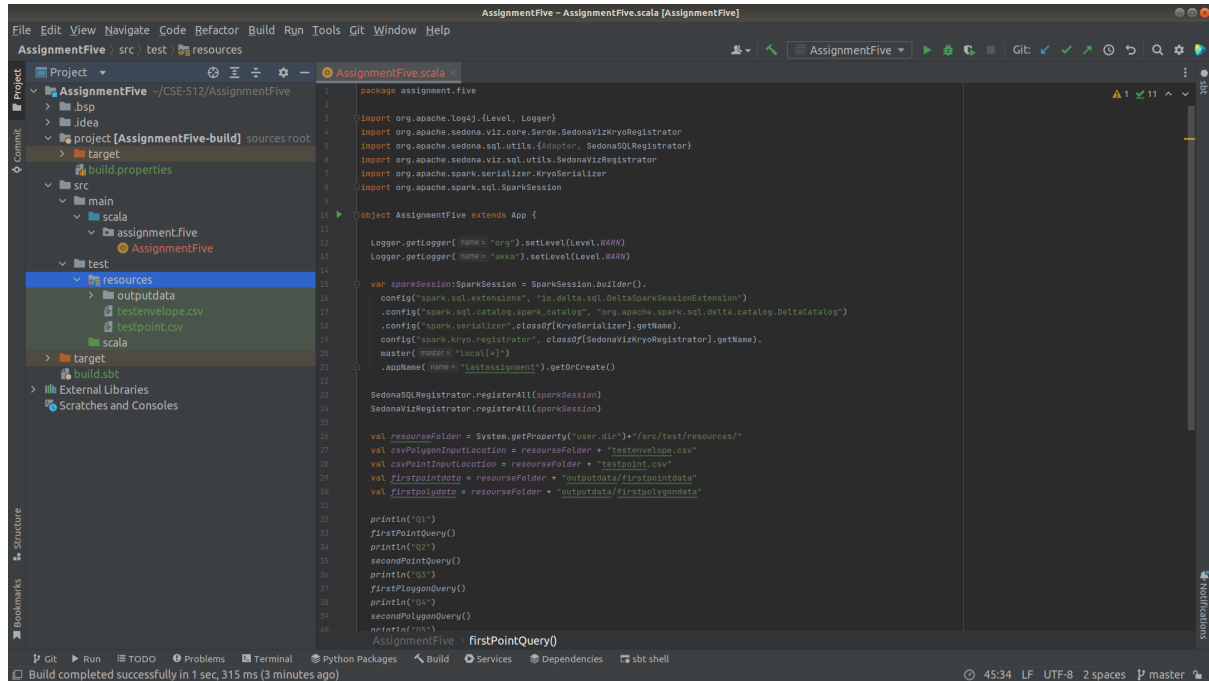


CSE 512 Assignment 5 (Delta Lake)

This assignment requires you to setup sbt project and fill up five sql queries. Steps for the assignment are as follows:

1. Download the provided Zip from canvas. You have to set up your IntelliJ IDE with it. This project includes the setup for [Apache Sedona](#) and [Delta Lake](#). It should look like as follows



2. Open the Scala file and write following five queries in given order

Input Data: The point dataset (testpoint.csv) consists of coordinates (x,y) of various points while the polygon dataset (testenvelope.csv) consists of coordinates of two diagonal points (x1, y1, x2, y2) of rectangles.

2.1 Read the given testpoint.csv file in csv format. Write the points whose two attributes are all larger than 500 in delta format and save it named firstpointdata. I.e. firstPointQuery()

2.2 Read the firstpointdata in delta format. Print the total count of the points. I.e. secondPointQuery()

2.3 Read the given testenvelope.csv in csv format. Write the polygons whose four attributes are all larger than 900 in delta format and save it named firstpolydata. I.e. firstPolygonQuery()

2.4 Read the firstpolydata in delta format. Print the total count of the polygon. I.e. secondPolygonQuery()

2.5 Read the firstpointdata and firstpolydata in delta format. Find the total count for point pairs where the distance between the points within a pair is less than 2, and both points are within some rectangles from firstpolydata. Note the two points do not need to be in the same rectangle. I.e. JoinQuery()

3. You have to submit only the AssignmentFive.scala file to canvas with complete code.