# Spatial Data Science & Engineering

## Assignment 3

### Maximum points Possible – 10

### **Required Tasks**

**Set Up Working Environment:**

Setting up working environment is similar to the setup for Part-1 of Assignment-1. If you successfully did the setup for Assignment-1, you are done with the setup. If you didn't, setup Java, Hadoop, and Spark following the instructions provided with Assignment-1. In order to work with an IDEA such as Intellij IDEA, you may need to add the scala plugin. Install 'sbt' on your computer if you didn't for Assignment-1. Assignment-1 also provides instructions on how to run a project and test with input datasets.

**Tasks:**

1. Search one or more suitable datasets to test the ST function you selected in the Google Excel Sheet. The datasets can be either geometry data or raster data based on the selected ST function. You can use the datasets provided with Assignment-1 and Assignment-2 if they are suitable to test your ST function. Place the datasets under data/inputs folder of the given Spatial-UDF project. Some other dataset links are available at the end of this PDF file.

2. Open the Spark Scala project Spatial-UDF. Write a user defined function which performs the operation you selected in the Google Excel Sheet. Input parameters and output of the function are similar to its counterpart on PostGIS. Register the user defined function with the name you selected in the Excel Sheet. Load the datasets you selected to test your function. Create temporary views for the datasets. Write SQL queries using the registered ST function on the temporary views. Write the SQL output under the data/outputs folder. The written output must have the values computed by the registered ST function. Writing other column values in the output file are optional.

3. You need to run the selected ST function on PostGIS using the same datasets. Write the outputs retuned by PostGIS similarly you did for the Spark Scala project. Place the PostGIS output file under data/true-outputs folder.

4. Write a small report on how to run your project. If we need to download the dataset for testing, provide the links to the datasets. You should give proper, clear, and unambiguous instructions to run your project. If we cannot run your project, you won't get points.

**Submission Instructions:**

- You should submit a .zip file containing the full Spark Scala project and the report. If the datasets required to test your function are small (less than few MBs), you can put the datasets under data/inputs folder. Otherwise, provide links to the datasets in the report. Put the PostGIS output file under data/true-outputs folder. If the PostGIS output file is missing, you won't receive points. Delete the output file from data/outputs. If you submit the output generated by your written function, you will lose points.

- You need to make sure your project code can be compiled by entering *sbt clean assembly* command under the root folder where build.sbt file is located.

Extra Dataset Links:

- https://geodacenter.github.io/data-and-lab/
- https://drive.google.com/file/d/1X9VaxC2AYC-puRy7PkTCPIpWPw1t0P7O/view?usp=sharing
  You need to access this dataset with ASU email. This dataset is a point dataset corresponding to Taxi Zones Polygon shape file used in Assignment-1. If someone needs both points and polygons/multipolygons in separate datasets, this point dataset with taxi zones shape file is a good choice.
- https://drive.google.com/file/d/14L9XSnWVgZWaG8MS2PST7vT2NRVZnSPg/view?usp=sharing
  You need to access this dataset with ASU email. This is a raster dataset.