

# Lab 12: Analysis of New York Taxi pickup density in various traffic analysis zones using Spark

## 1. Outline

In this lab, you will perform the task for generating the New York taxi pickup density map based on the traffic analysis zones (TAZ) and visualize the results in ArcGIS. The final result should be sent to Junjun ([jyn@illinois.edu](mailto:jyn@illinois.edu)) as assignment 2.

## 2. Materials

The data and scripts are stored in: `/gpfs_scratch/geog479/assignment2`

Before copying this folder to your home directory:

Delete the same folder in the your home directory from previous lecture:

```
>> rm -r ~/assignment2
```

And then copy

```
>> cp -r /gpfs_scratch/geog479/assignment2 ~/
```

## 3. Tasks

### Task 1:

- The taxi data is ready in `ny_taxi_1.csv`, containing over 14 million taxi records during January, 2013
- The shapefile of the TAZ of New York city is located in the same folder named “shape”
  - You can copy (scp) the `osm_shp_NY.zip` to your local desktop and open it with ArcGIS
- For this lab, we can use small portion of the datasets, say, 10,000
  - `head -n 10000 ny_taxi_1.csv > example.csv`
- Remember from previous lab, as we need to determine which TAZ the pickup happened (basically, a point-in-polygon process). Therefore you need to create a spatial index for the shapefile to speed things up
  - `python buildTree.py shape/NY_TAZ.shp quadTree.txt`
- Now, let's go to the main Spark script: `NY_Taxi_Spark.py`
- To complete the script, you can refer to the similar one in `/gpfs_scratch/geog479/lecture11`
  - Can you recall on how to prepare the (key, values) for summarizing the number of passengers in a zone and the number of taxi in a zone?
  - How to calculate the pickup density in every zone?
- When you are ready, execute the program
  - `spark-submit --master yarn-client --executor-memory 10g NY_Taxi_Spark.py`

### Task 2:

- Visualize the results in ArcGIS
- Expand your data to the whole dataset
- Modify the code for enabling criteria of constraining time and space range