**Distributed Tweet Location Project**

Goals:

1. Load tweet files into separate SQLite databases in 1-day batches
   1. Database will be named **<batch\_date>.db**
   2. Include user\_id, tweet\_text, tweet\_location, tweet\_date, fips as table columns
   3. This table will be called **tweets**
2. Get the modal location for each 1-day batch of users
3. Check for which users the modal location changes

Steps:

1. Import tweets to database with **import\_tweets\_to\_db.py**
   1. Chunk tweet file by day
   2. Loop through day chunks
      1. Loop through files in chunk
      2. Loop through tweets in file
         1. Insert tweets into database
2. Save unique users to tweets database with **get\_users\_per\_db.py**
3. Loop through unique users in each database with **get\_user\_fips\_per\_db.py**
   1. Save user\_id and fips in table called users\_fips
4. Loop through records in each users\_fips table and create a new database **aggregated\_data.db**
   1. Insert unique user\_id + fips records into new table with **join\_users\_fips.py**
5. Get potential movers with **get\_potential\_movers.py**
   1. Loop through all users in **aggregated\_data.db** table users\_fips
   2. Fetch COUNT(user\_id)
      1. If count > 1 this is a potential mover
6. Get actual movers with **get\_actual\_movers.py**
   1. Get unique potential movers
   2. Chunk potential movers into batches of 10000
   3. Loop through user chunks
      1. Loop through each user in each chunk
      2. Check if user has moved and if so get all moves
      3. If user is mover, insert user\_id, fips1, date 1, fips 2, date 2 into **actual\_movers.db**
7. Combine (fips1, date1) and (fips2, date2) into one column
   1. Sort and chunk by user