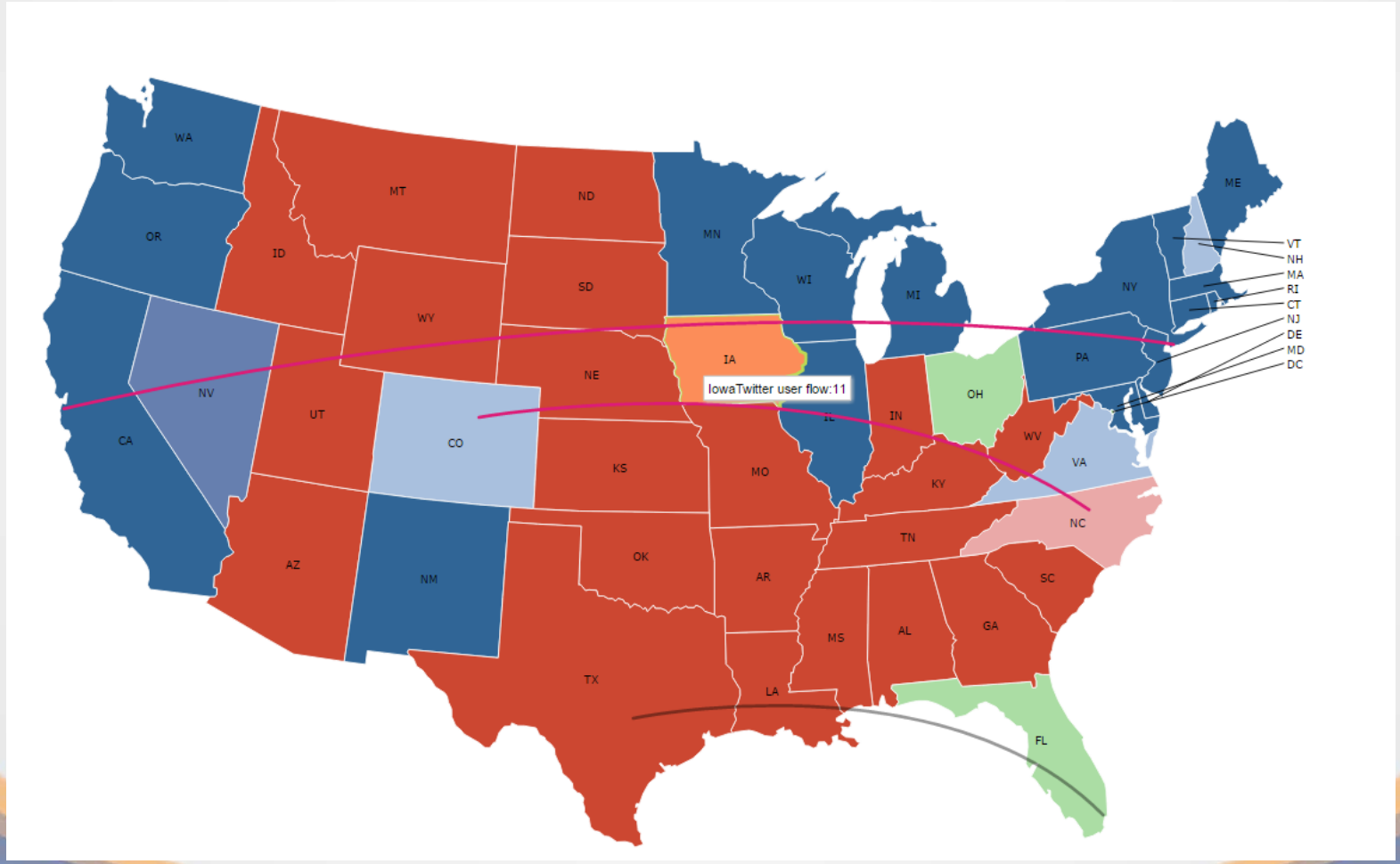
Interactive Visualization of Large-scale Movement Data using Apache Spark

Junjun Yin

CyberGIS Center for Advanced Digital and Spatial Studies
Department of Geography and Geographic Information Science
National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign

April 12, 2016

Today's main objective:



How to get the visualization

- It is based on DataMaps: http://datamaps.github.io/
- An extended library based on D3.js

- How to customize the existing source code base for your own use
  - Create your own map layers with TopoJSON (https://github.com/mbostock/topojson/wiki)
  - Choose one of the existing template to get started

- Setup an http server to host the web page
  - E.g., python -m SimpleHTTPServer 8000
  - Note, to use the lab computer, you need to use Python that comes along with ArcGIS

Let's go through the script

- The data is located in ROGER: /gpfs_scratch/geog479/lecture11
- The web page functional based on JavaScript
- The TopoJSON is modified to exclude Alaska and Hawaii
- Pay attention to how to set the color values and how to draw arcs
- Question: How to calculate the origin and destination coordinates?

How Spark can help prepare the data?

- For this case, we are utilizing Twitter data to generate the flows
  - You can apply it to the Taxi data, or any other movement data that track people's movement from A to B

How to calculate the volume of tweets in each state?

How to summarize the movement flux among different states?

Calculate the volume of tweets in each state

- Before we get started, what is your plan?
- Let's examine the details in spark_pinP.py

Summarize the movement flux

- Now we know the trajectory of each user travelling from one state to another
  - How do we summarize the flows among states (in a MapReduce way)
- How do we incorporate such results to (interactive) visualization?