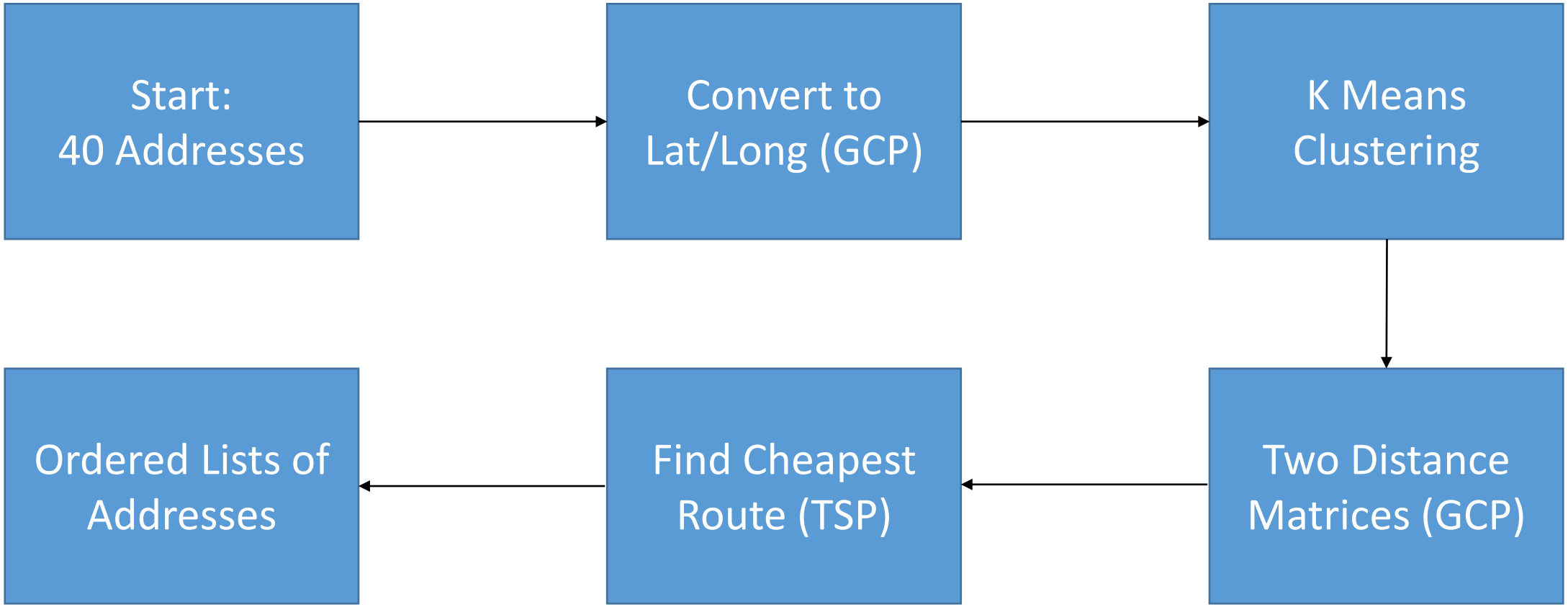
Travelling Landscaper

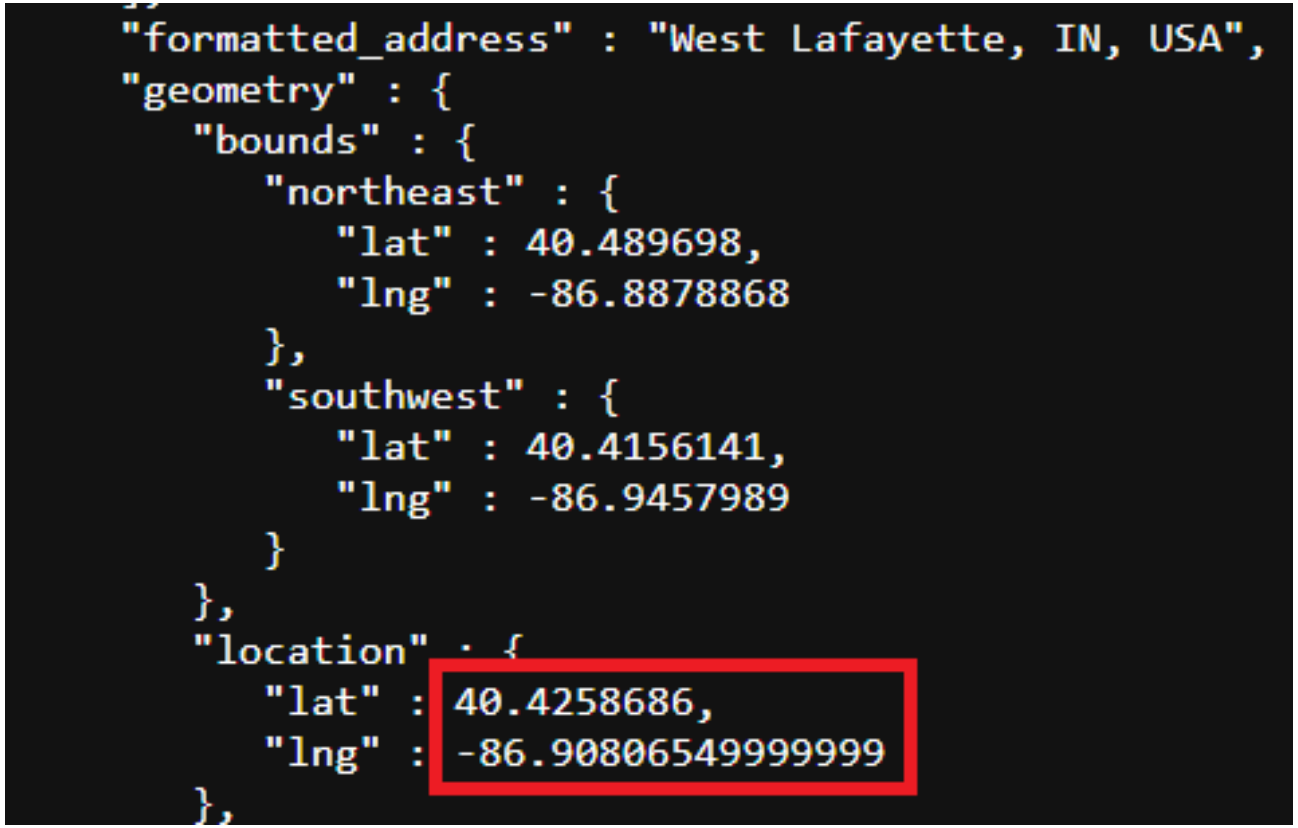
Ben Meshanko

Software Pipeline Overview



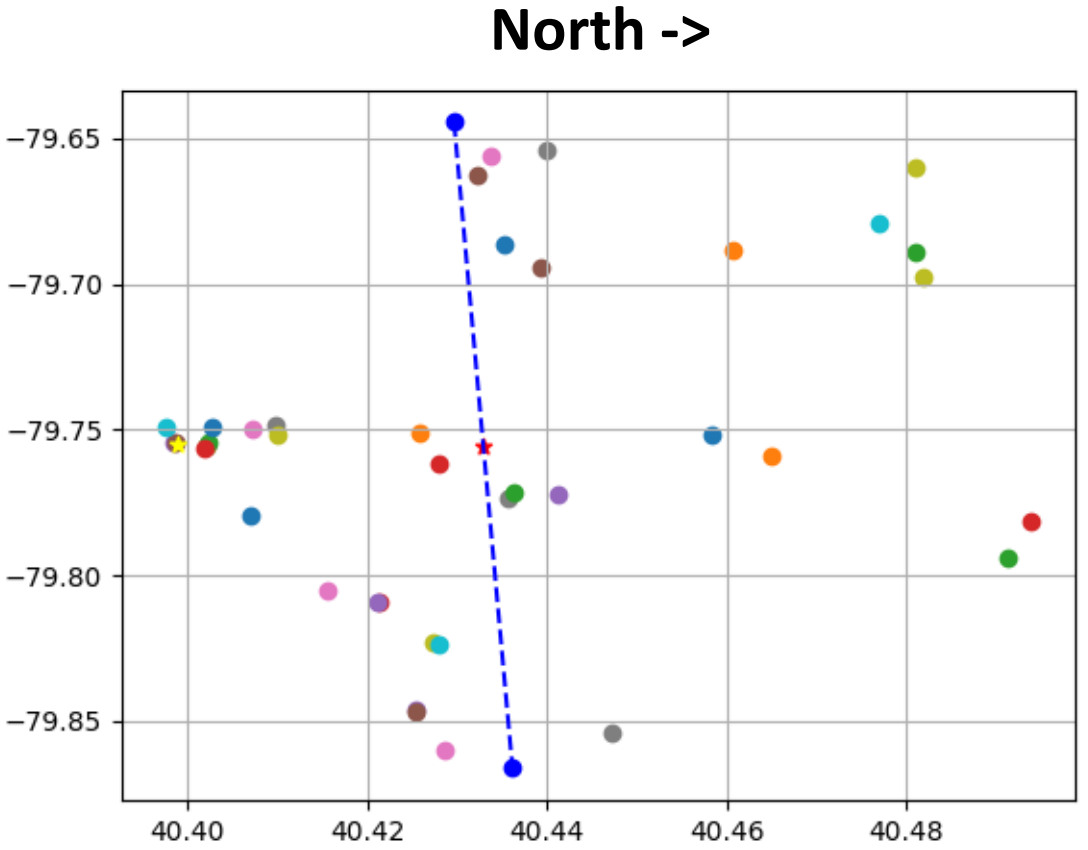
Conversion to Coordinates

- Geocoding API
- Address -> Lat/Long
- Python requests library



K Means Clustering

- Convert Lat/Long to nparray
- Ran Clustering Algorithm (G4G)
- Python matplotlib library
- Used center to determine two clusters, 40 -> 18/22



Note: The Earth is **not** flat, but the points are close enough to not skew

Distance Matrix from Clusters

- Distance Matrix API
- Parse “duration” to int representing minutes
- Added in origin, clusters are now 19/23

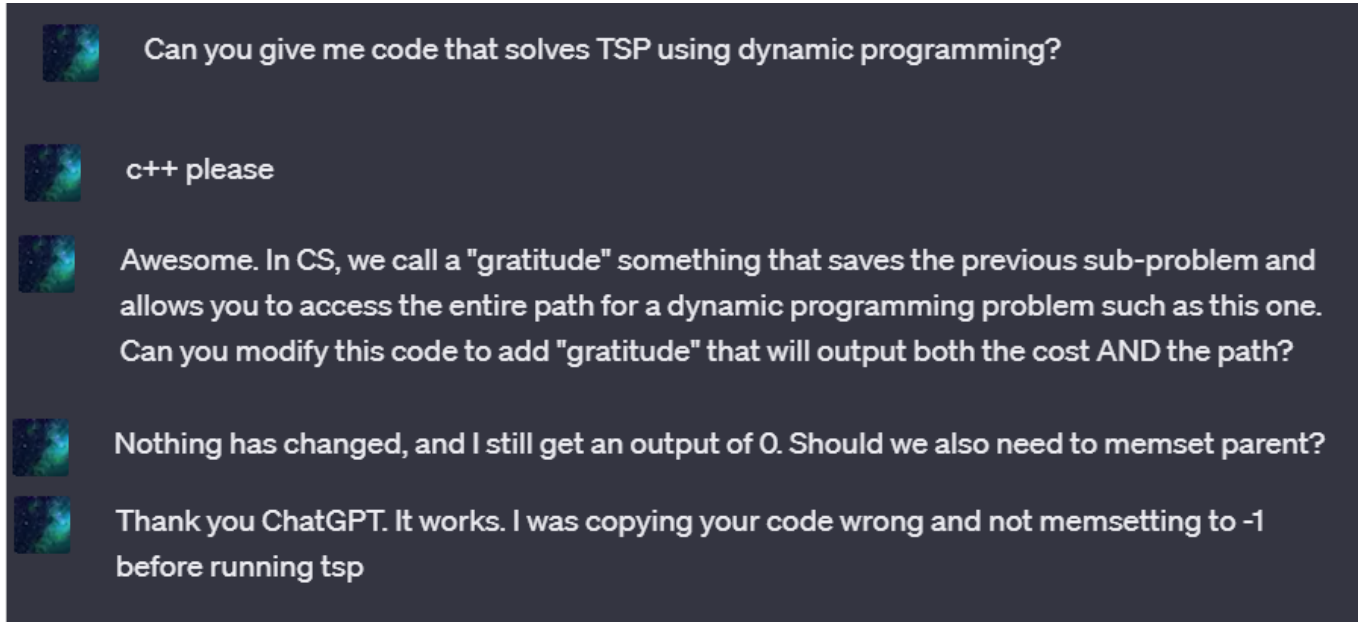

```
"destination_addresses" : [ "Bloomington, IN, USA" ],
"origin_addresses" : [ "West Lafayette, IN, USA" ],
"rows" : [
   {
      "elements" : [
         {
            "distance" : {
               "text" : "114 mi",
               "value" : 183735
            },
            "duration" : {
               "text" : "2 hours 5 mins",
               "value" : 7471
            },
            "status" : "OK"
         }
      ]
   }
],
"status" : "OK"
```

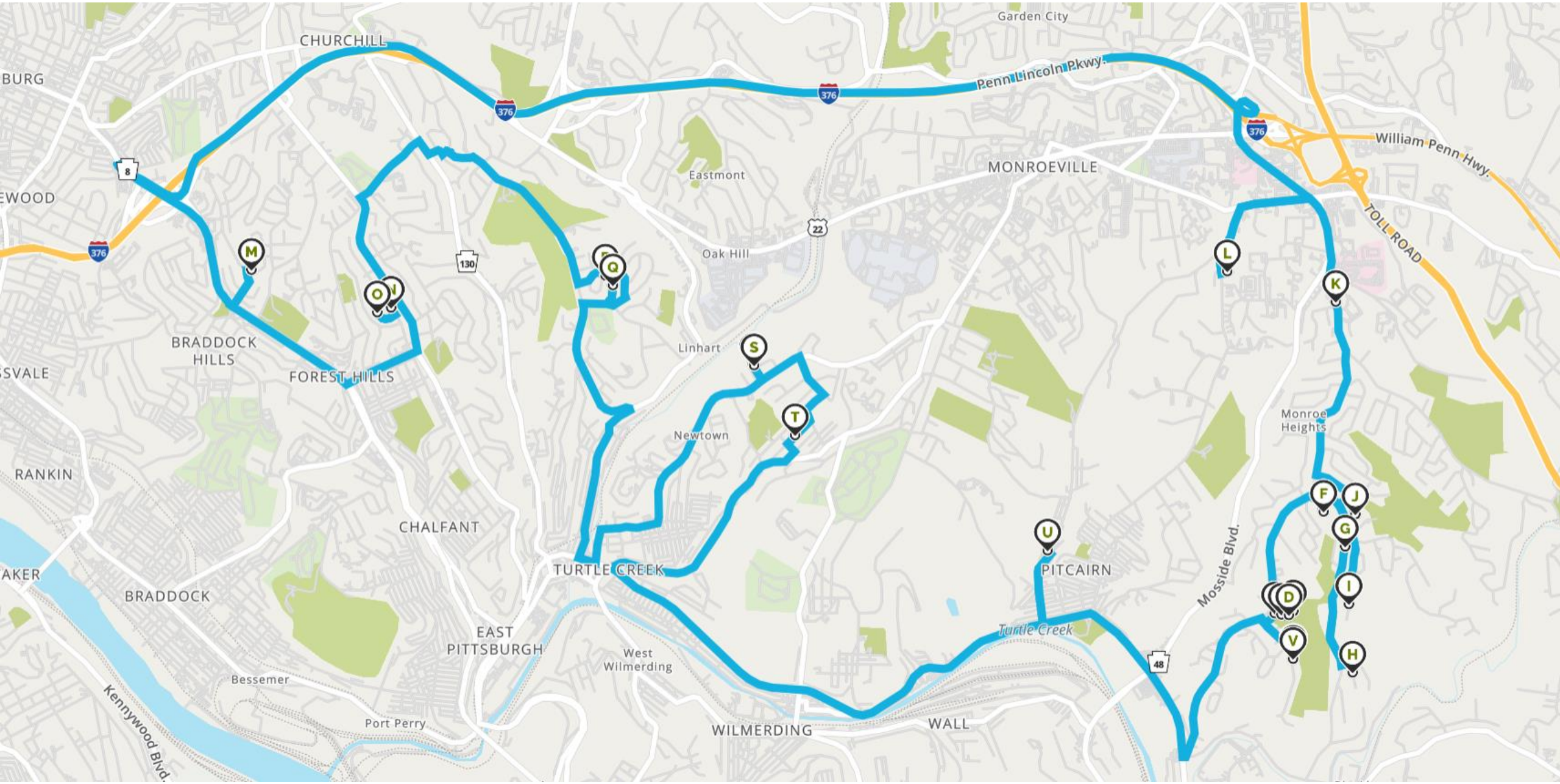
Computing Shortest Path with TSP

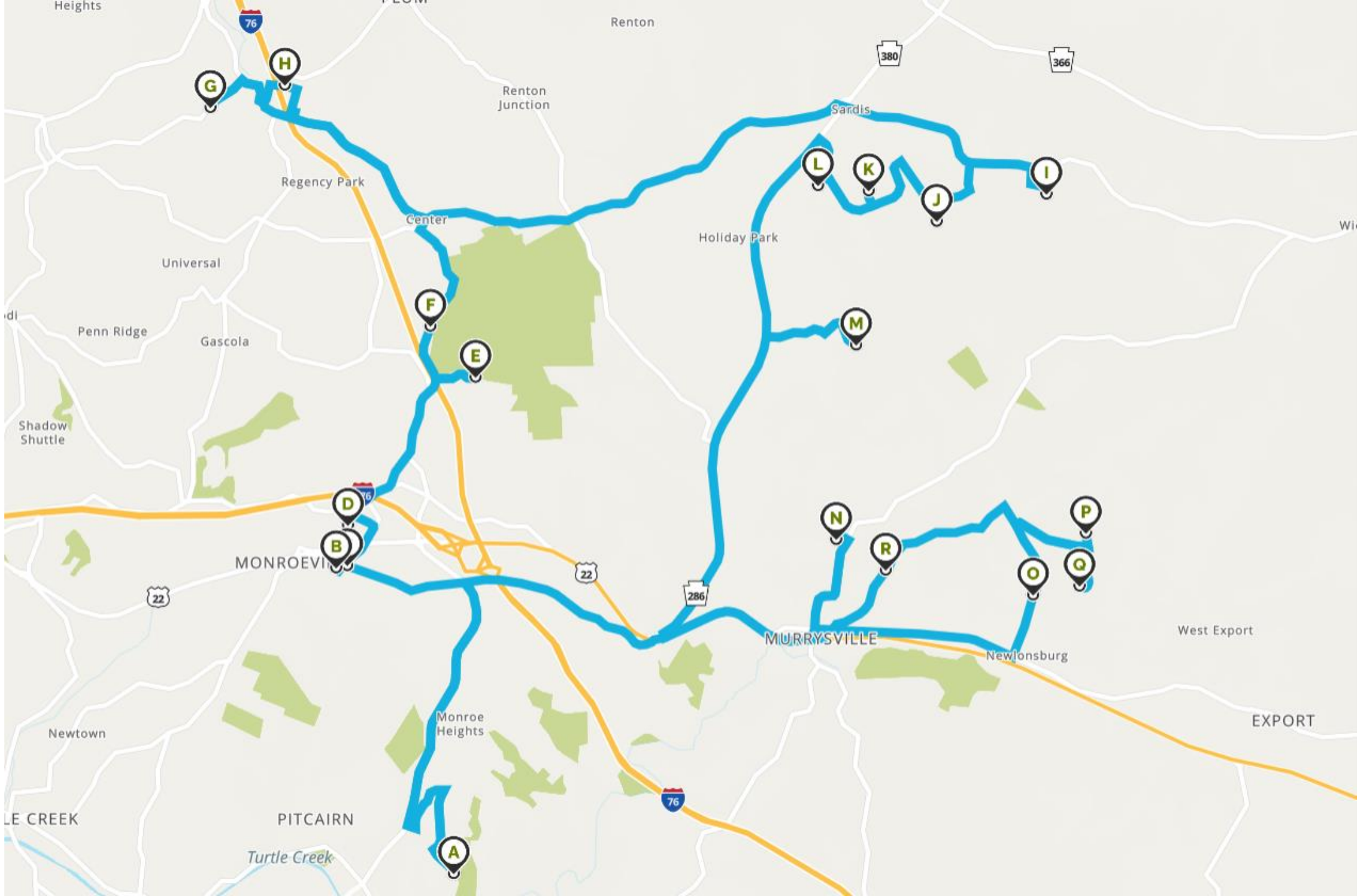
- Tried O(n!) implementation in Python, quickly gave up
  - 23! = 2.58e22
  - C++ is far faster anyways

- Runtime: $O(n^2 2^n)$.

- Outputs abstract 0-indexed graph path



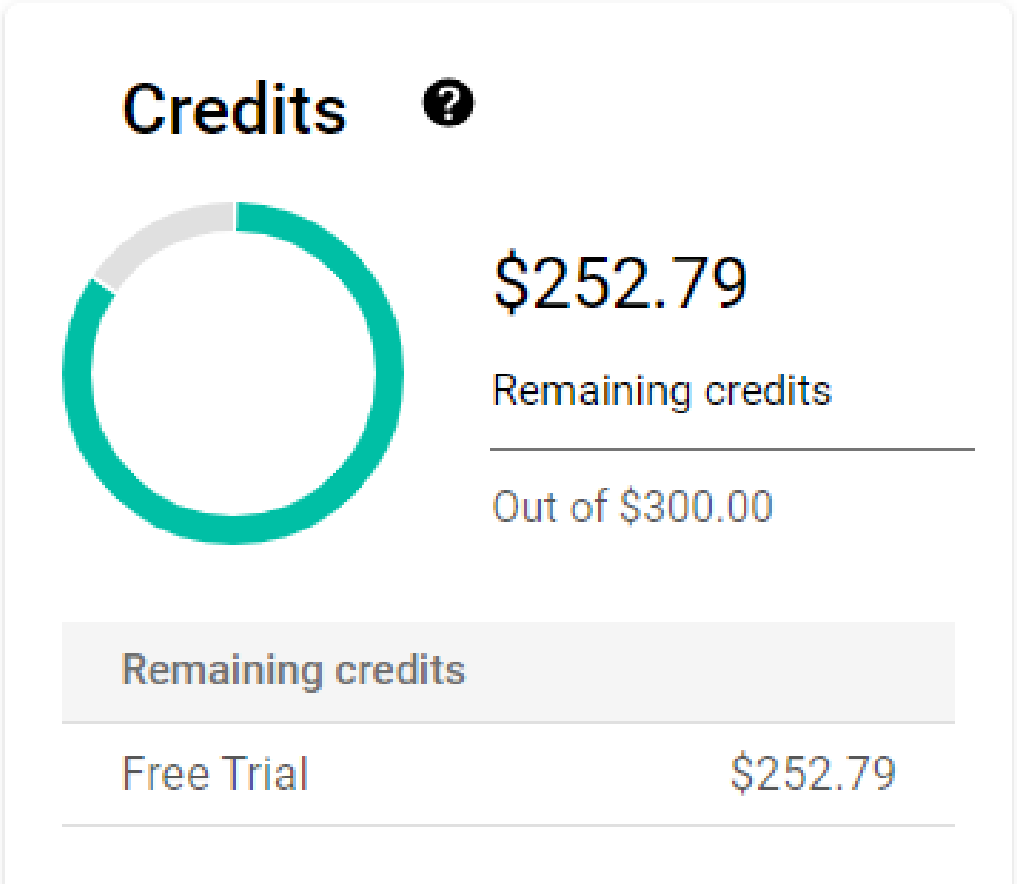

CHURCHILL
Garden City
Penn Lincoln Pkwy.
William Penn Hwy.
TOLL ROAD
MONROEVILLE
Eastmont
Oak Hill
Linhart
Newtown
BRADDOCK HILLS
FOREST HILLS
RANKIN
CHALFANT
TURTLE CREEK
PITCAIRN
Monroe Heights
Mosside Blvd.
BRADDOCK
EAST PITTSBURGH
West Wilmerding
Turtle Creek
Bessemer
Port Perry
WILMERDING
WALL
Kennywood Blvd.
M
O
N
Q
S
T
U
L
K
F
J
G
I
D
V
H
376
8
22
130
48


Heights
Renton
Renton Junction
Sardis
Regency Park
Center
Holiday Park
Universal
Penn Ridge
Gascola
Shadow Shuttle
MONROEVI
MURRYSVILLE
Newlonsburg
West Export
EXPORT
Monroe Heights
Newtown
PITCAIRN
Turtle Creek
76
22
286
380
366
A
B
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R

What I Learned/Struggled With

- GCP Maps API is pretty incredible, but
- Python is sub-optimal for intensive computation
- Larger Sample Size -> Move to Cloud
  - GPU Optimizations
- One last thing…

Remember this question from HW3?

20. You have been hired by a company that is moving an application that processes over a hundred
    Terabytes of data to the cloud. None of the software engineers has any experience with Hadoop,
    and they all need to learn how to use it. Your manager is incredibly worried about the
    ``training cost'' (i.e., the cloud charges that will be incurred while the software engineers
    learn about Hadoop and debug various pieces of software. What do you recommend to your
    manage as a way to avoid the training expenses?

I also depleted the $50 educational credit…