**CS 575 Parallel Programming**

**Project #3**

**K-means: A Real Application Parallel Challenge**

**Name**: Rachel Xing (#934229189)

**Email Address**: xingru@oregonstate.edu

**Description**: This project uses OpenMP to parallelize the K-means algorithm that automatically re-divides the United States (excluding Alaska and Hawaii) by re-assigning capitals among the 331 largest cities. The number of new capitals is user-defined, and each new capital is the centroid of the cities among its respective new state region. The algorithm runs for 100 iterations to converge on the new capital locations. The portion of each iteration that re-assigns cities to capitals based on the updated locations from the previous iteration is parallelized. Then, after identifying the locations of the new capitals, iterate through the list of cities for each capital to find the closest city as the new capital (extra credit). Performance is measured in terms of MegaCityCapitals computed per second. The program is tested with thread numbers of 1, 2, 4, 6, 8, and capital numbers of 2, 3, 4, 5, 10, 15, 20, 30, 40, 50.

**Commentary**:

1. **Tell what machine you ran this on**

   The machine I ran the program on is the **OSU Rabbit server**, which is equipped with Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz.

2. **Tell what operating system you were using**

   The operating system used by the OSU Rabbit server is **CentOS Linux 7 (Core)**, which is derived from Red Hat(R) Enterprise Linux(R).

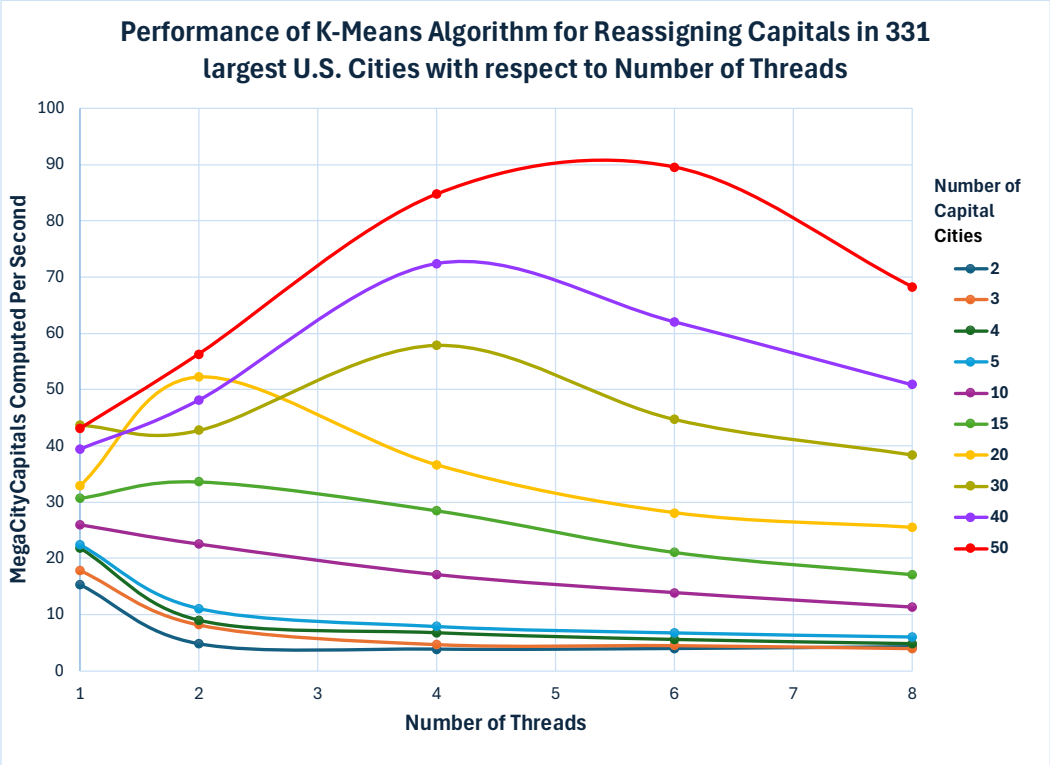
3. **Tell what compiler you used**

   The compiler I used is the **GNU C++ Compiler** (OpenMP enabled).

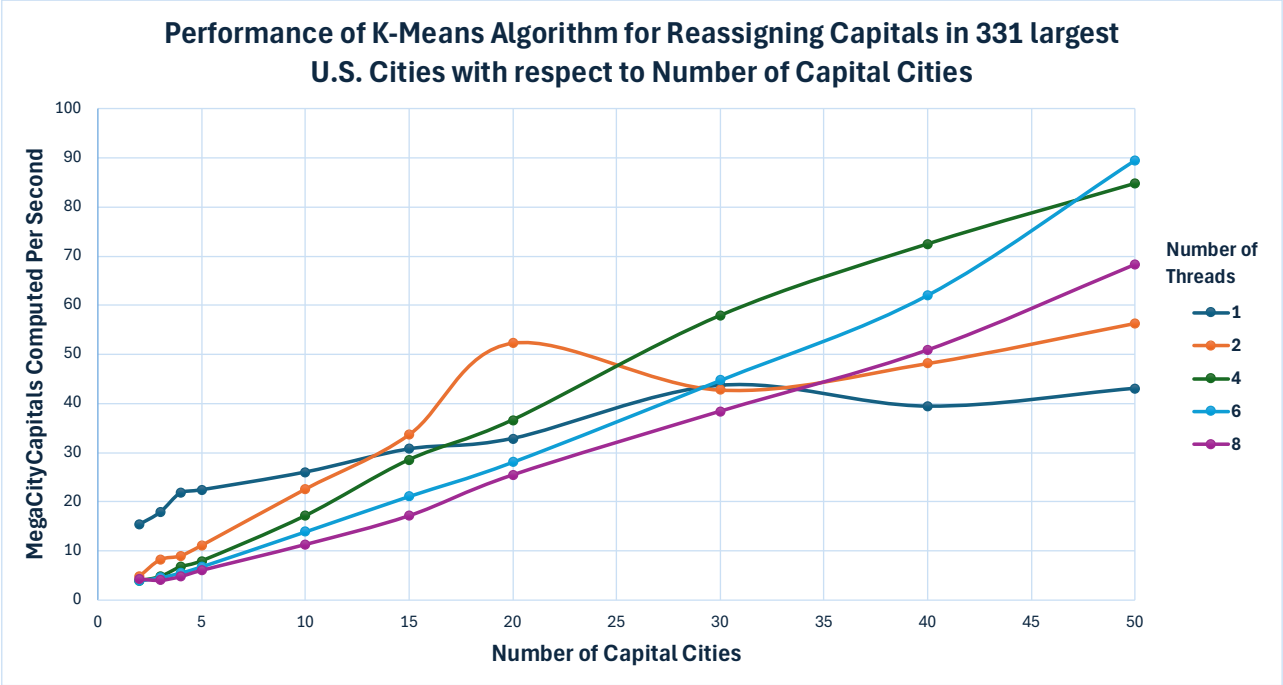
4. **Include the rectangular table of performance data.**

| Performance (MegaCityCapitals/sec) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Capitals # / Threads # | 2 | 3 | 4 | 5 | 10 | 15 | 20 | 30 | 40 | 50 |
| 1 | 15.374 | 17.839 | 21.856 | 22.418 | 26.026 | 30.729 | 32.87 | 43.669 | 39.428 | 43.064 |
| 2 | 4.809 | 8.191 | 8.968 | 11.1 | 22.571 | 33.632 | 52.301 | 42.752 | 48.155 | 56.325 |
| 4 | 3.85 | 4.715 | 6.725 | 7.898 | 17.109 | 28.502 | 36.621 | 57.865 | 72.429 | 84.746 |
| 6 | 3.959 | 4.535 | 5.527 | 6.762 | 13.889 | 21.096 | 28.075 | 44.721 | 62.064 | 89.548 |
| 8 | 4.285 | 4.004 | 4.772 | 6.038 | 11.293 | 17.149 | 25.455 | 38.409 | 50.871 | 68.265 |

5. **Include a graph of performance vs. NUMT with the colored curves being NUMCAPITALS.**




6. **Include a graph of performance vs. NUMCAPITALS cities with the colored curves being NUMT.**



7. **Tell us what you discovered by doing this. What patterns are you seeing in the graphs?**

    - Because the portion of the K-means algorithm that reassigns cities to their nearest capital is parallelized, and each city computes and compares its distance to all capitals, the computational workload per iteration increases with the product of the number of cities and the number of capitals. Since the number of cities is fixed (i.e., 331) in this problem, **the problem scale is determined by the number of capitals**.
    - As shown in the chart for problem 6, **increasing the problem scale (the number of capitals) leads to improved performance for all thread numbers greater than one**. With NUMT = 1, performance reaches a plateau at around 40 MegaCityCapitals per second once NUMCAPITALS exceeds 20. This trend aligns well with Gustafson's law.
    - **For smaller problem scales (with capital numbers of 2, 3, 4, 5, or 10), we can observe from the above two charts that peak performance is achieved by a single thread (NUMT = 1), and increasing the number of threads degrades the performance**. That can be explained by Amdahl's Law that when the parallelizable portion of the task is too small for parallelization, the overheads introduced by parallelization, such as thread management and potential false sharing, can outweigh its gain. As a result, a single thread is optimal for such small computation workloads.
    - **For medium to large problem scales, the performance improves with more threads until reaching a "sweet spot".** We can observe from the above two charts that the peak performance occurs with 2 threads when the number of capitals is 15 or 20, with 4 threads for 30 and 40 capitals, and with 6 threads for 50 capitals. That is because as the problem scale increases, the parallelizable portion of the task grows, allowing the program to have a net performance gain from parallelization. However, beyond a certain number of threads, the overhead dominates again and leads to performance degradation. Thus, such a trade-off between parallelization gain and overheads creates a "sweet spot".
    - If the problem scale continues to increase, we can then predict that the optimal number of threads will also increase. For a certain large number of capitals, a big enough parallelizable portion of the task will make 8 threads optimal. Thus, we conclude that **for the same problem, the optimal number of threads increases with the problem size, and there is always a "sweet spot" for performance**.

8. **(Extra Credit) When you are done computing the final longitudes-latitudes of the NUMCAPITALS, go through the list of cities and print the name of the city that is closest to each capital's longitude-latitude.**

    **Format for Outputs:** `[New capital index: Longitude, Latitude, Closest city name]`

1. When NUMCAPITALS = 2:

```
  0:     85.82 ,    36.04 , Murfreesboro,TN
  1:    116.27 ,    37.16 , LasVegas,NV
```

2. When NUMCAPITALS = 3:

```
  0:     77.42 ,    39.58 , Washington,DC
  1:     90.98 ,    33.99 , Memphis,TN
  2:    116.71 ,    37.28 , LasVegas,NV
```

3. When NUMCAPITALS = 4:

```
  0:     74.24 ,    40.72 , Newark,NJ
  1:     83.91 ,    35.14 , Knoxville,TN
  2:     98.01 ,    35.00 , OklahomaCity,OK
  3:    118.27 ,    37.15 , Visalia,CA
```

4. When NUMCAPITALS = 5:

```
  0:     74.24 ,    40.72 , Newark,NJ
  1:    108.10 ,    36.56 , RioRancho,NM
  2:     82.85 ,    34.20 , Athens,GA
  3:     94.64 ,    35.37 , BrokenArrow,OK
  4:    119.45 ,    37.39 , Clovis,CA
```

5. When NUMCAPITALS = 10:

```
  0:     74.10 ,    40.73 , JerseyCity,NJ
  1:     92.01 ,    41.15 , CedarRapids,IA
  2:    111.39 ,    33.45 , Mesa,AZ
  3:     81.58 ,    28.82 , Orlando,FL
  4:    120.86 ,    45.93 , Gresham,OR
  5:    117.77 ,    34.11 , Pomona,CA
  6:     96.70 ,    31.77 , Waco,TX
  7:    106.77 ,    39.74 , Boulder,CO
  8:     83.57 ,    37.95 , Lexington,KY
  9:    121.65 ,    37.98 , Antioch,CA
```

6. When NUMCAPITALS = 15:

```
  0:     73.03 ,    41.52 , Waterbury,CT
  1:    112.53 ,    35.42 , Peoria,AZ
  2:    119.66 ,    36.04 , Visalia,CA
  3:     96.97 ,    31.62 , Waco,TX
  4:    120.86 ,    45.93 , Gresham,OR
  5:     80.99 ,    27.33 , PortSt.Lucie,FL
  6:     86.47 ,    33.49 , Birmingham,AL
  7:     78.58 ,    36.00 , Raleigh,NC
  8:     82.22 ,    41.80 , Cleveland,OH
  9:     87.51 ,    40.93 , Joliet,IL
 10:     94.56 ,    40.89 , DesMoines,IA
 11:    105.34 ,    38.45 , ColoradoSprings,CO
```

```
12:    118.41 ,    34.17 , Burbank,CA
13:    117.47 ,    33.72 , Corona,CA
14:    121.80 ,    38.07 , Antioch,CA
```

7. When NUMCAPITALS = 20:

```
 0:     71.89 ,    42.10 , Worcester,MA
 1:    121.03 ,    47.58 , Bellevue,WA
 2:    120.46 ,    39.63 , Reno,NV
 3:     96.22 ,    29.13 , SugarLand,TX
 4:     88.08 ,    41.50 , Joliet,IL
 5:     78.79 ,    35.47 , Cary,NC
 6:    119.66 ,    36.04 , Visalia,CA
 7:     97.41 ,    33.42 , Denton,TX
 8:     85.75 ,    32.61 , Montgomery,AL
 9:    117.65 ,    34.02 , Ontario,CA
10:     80.98 ,    27.10 , PortSt.Lucie,FL
11:    106.77 ,    39.74 , Boulder,CO
12:     75.05 ,    40.63 , Allentown,PA
13:     82.76 ,    41.65 , Dearborn,MI
14:    121.75 ,    38.23 , Antioch,CA
15:    111.39 ,    33.45 , Mesa,AZ
16:     86.22 ,    36.62 , Nashville,TN
17:    120.73 ,    44.60 , Bend,OR
18:     94.40 ,    41.06 , DesMoines,IA
19:    122.18 ,    37.54 , Hayward,CA
```

8. When NUMCAPITALS = 30:

```
 0:     74.50 ,    40.78 , Woodbridge,NJ
 1:    121.67 ,    37.41 , SanJose,CA
 2:     78.82 ,    42.17 , Buffalo,NY
 3:    120.93 ,    39.10 , Roseville,CA
 4:     94.48 ,    44.87 , Minneapolis,MN
 5:    119.60 ,    36.65 , Clovis,CA
 6:    105.18 ,    39.06 , ColoradoSprings,CO
 7:     89.23 ,    42.98 , Madison,WI
 8:    113.17 ,    42.36 , SaltLakeCity,UT
 9:     80.98 ,    27.10 , PortSt.Lucie,FL
10:     73.10 ,    41.67 , Waterbury,CT
11:     86.03 ,    32.56 , Montgomery,AL
12:     76.62 ,    37.67 , NewportNews,VA
13:    122.87 ,    38.09 , SantaRosa,CA
14:    111.39 ,    33.45 , Mesa,AZ
15:     94.79 ,    38.59 , Olathe,KS
16:     88.97 ,    41.07 , Peoria,IL
17:    118.02 ,    34.00 , ElMonte,CA
18:    116.93 ,    34.01 , MorenoValley,CA
19:     97.95 ,    32.87 , FortWorth,TX
```

```
20:    119.26 ,    34.53 , Ventura,CA
21:     79.93 ,    34.72 , Concord,NC
22:     86.39 ,    36.37 , Nashville,TN
23:     71.21 ,    42.29 , Cambridge,MA
24:    121.03 ,    47.58 , Bellevue,WA
25:     94.57 ,    30.38 , Beaumont,TX
26:    122.58 ,    45.04 , Salem,OR
27:     98.17 ,    27.57 , CorpusChristi,TX
28:     84.14 ,    41.30 , Toledo,OH
29:    122.17 ,    37.92 , Berkeley,CA
```

9. When NUMCAPITALS = 40:

```
 0:     78.15 ,    42.39 , Buffalo,NY
 1:     95.65 ,    36.07 , BrokenArrow,OK
 2:    122.46 ,    37.89 , Richmond,CA
 3:    122.26 ,    47.55 , Renton,WA
 4:    119.60 ,    36.65 , Clovis,CA
 5:    117.67 ,    33.81 , Anaheim,CA
 6:     96.88 ,    32.97 , Carrollton,TX
 7:     97.76 ,    26.57 , Edinburg,TX
 8:     79.48 ,    35.57 , Greensboro,NC
 9:     84.23 ,    40.98 , FortWayne,IN
10:     99.49 ,    27.56 , Laredo,TX
11:     82.36 ,    27.69 , St.Petersburg,FL
12:    113.32 ,    38.49 , St.George,UT
13:     76.62 ,    37.67 , NewportNews,VA
14:     94.47 ,    41.29 , DesMoines,IA
15:     87.60 ,    34.06 , Tuscaloosa,AL
16:    110.66 ,    33.44 , Mesa,AZ
17:    122.21 ,    44.06 , Bend,OR
18:     80.29 ,    25.92 , Hialeah,FL
19:     97.61 ,    30.46 , RoundRock,TX
20:     80.32 ,    33.26 , NorthCharleston,SC
21:     88.76 ,    41.47 , Aurora,IL
22:     82.75 ,    30.16 , Gainesville,FL
23:     84.31 ,    33.96 , SandySprings,GA
24:     74.35 ,    40.58 , Woodbridge,NJ
25:     80.26 ,    27.01 , WestPalmBeach,FL
26:    101.58 ,    33.02 , Lubbock,TX
27:     93.92 ,    30.15 , Beaumont,TX
28:     80.96 ,    28.18 , Orlando,FL
29:     81.61 ,    32.68 , Savannah,GA
30:     73.10 ,    41.67 , Waterbury,CT
31:    105.21 ,    40.18 , Boulder,CO
32:    122.73 ,    45.43 , Portland,OR
33:    119.03 ,    34.54 , Ventura,CA
34:    120.46 ,    39.63 , Reno,NV
35:     80.20 ,    26.15 , FortLauderdale,FL
```

```
36:     71.13 ,    41.85 , Brockton,MA
37:    116.77 ,    45.22 , Meridian,ID
38:     71.24 ,    42.48 , Cambridge,MA
39:    121.72 ,    37.85 , Antioch,CA
```

10. When NUMCAPITALS = 50:

```
 0:     74.08 ,    40.68 , JerseyCity,NJ
 1:     98.10 ,    26.77 , Edinburg,TX
 2:     83.90 ,    39.64 , Dayton,OH
 3:     85.24 ,    33.60 , SouthFulton,GA
 4:     85.45 ,    36.84 , Louisville,KY
 5:    119.60 ,    36.65 , Clovis,CA
 6:     77.31 ,    36.38 , NewportNews,VA
 7:    119.04 ,    35.35 , Bakersfield,CA
 8:     83.74 ,    42.42 , AnnArbor,MI
 9:     82.42 ,    28.83 , Gainesville,FL
10:     80.43 ,    34.46 , Columbia,SC
11:    114.74 ,    36.40 , NorthLasVegas,NV
12:     85.94 ,    40.50 , Fishers,IN
13:    112.00 ,    33.37 , Tempe,AZ
14:    122.26 ,    47.55 , Renton,WA
15:    117.33 ,    34.47 , Victorville,CA
16:    117.44 ,    34.03 , JurupaValley,CA
17:    108.55 ,    45.79 , Billings,MT
18:    101.58 ,    33.02 , Lubbock,TX
19:     88.97 ,    41.99 , Rockford,IL
20:     81.06 ,    41.00 , Akron,OH
21:     91.08 ,    31.55 , BatonRouge,LA
22:     88.77 ,    37.08 , Clarksville,TN
23:    104.93 ,    39.71 , Denver,CO
24:    122.17 ,    37.92 , Berkeley,CA
25:     71.21 ,    42.29 , Cambridge,MA
26:    117.13 ,    32.95 , SanDiego,CA
27:     95.18 ,    29.78 , Pasadena,TX
28:     77.54 ,    43.03 , Rochester,NY
29:     76.90 ,    39.01 , Washington,DC
30:     94.34 ,    44.32 , Minneapolis,MN
31:     75.31 ,    40.30 , Philadelphia,PA
32:    116.40 ,    43.60 , Meridian,ID
33:    121.67 ,    37.41 , SanJose,CA
34:     97.61 ,    30.46 , RoundRock,TX
35:     95.04 ,    38.65 , Olathe,KS
36:    120.93 ,    39.10 , Roseville,CA
37:    111.98 ,    40.69 , WestValleyCity,UT
38:     73.03 ,    41.77 , Waterbury,CT
39:    111.65 ,    40.25 , Provo,UT
40:    117.17 ,    33.58 , Murrieta,CA
41:    117.33 ,    47.67 , Spokane,WA
```

```
42:    119.63 ,    34.47 , Ventura,CA
43:    106.64 ,    33.64 , LasCruces,NM
44:     80.40 ,    26.36 , CoralSprings,FL
45:    117.93 ,    33.85 , Fullerton,CA
46:    118.38 ,    34.23 , Burbank,CA
47:     96.94 ,    33.22 , Frisco,TX
48:    122.58 ,    45.04 , Salem,OR
49:    122.87 ,    38.09 , SantaRosa,CA
```

[*Additional Graph*: 3D visualization of performance with respect to the number of threads and capital cities]