[HPC Project]

[K-mean Algorithm]

[Mohamed Kamel Sayed] | [Omar Sayed Mostafa] | [31-12-2016]

Structure of code

The code is divide into two main parts :

1. Sequential
2. Parallel using **OpenMp**

Steps of Algorithm For both Parts

1. Read test cases from file (sequential in two parts) .
2. Determine the number of Clusters Required **K.**
3. Choose first K number from input and make it initial means and calculate the time when it is in sequential and parallel and compare performance .
4. While loop start from 0 and stop at the number of epoch (number of epoch =20) .
5. In this while loop we find the closest mean to the element by calculating absolute difference for each element .
6. Assign the cluster number to this element.
7. We do this operation in all elements which it was read from file.
8. In update step we compute new mean value of cluster .
9. New mean value = the average of all elements in cluster, then Update the value of mean and repeat all steps again .

Test cases Length

1. 10
2. 10000
3. 1000000
4. 10000000

Performance

We run all test cases with number of clusters = 40 **Except** the first = 3 clusters.

Test cases	10	10000	100000	1000000	10000000
Read from file in sequential and parallel	0 s	0.075 S	0.687 S	6.454 S	65.514 S
Time in Sequential	0 s	0.333 S	3.359 S	35.145 S	355.326 S
Time in Parallel	0 s	0.146 S	1.227 S	12.182 S	159.924 S

Screenshot Test case 1000 000 with 100 cluster in sequential and parallel