CS-5379: Parallel Processing

Spring Semester 2017
Department of Computer Science
Texas Tech University

Assignment-03

Submitted By

Amitsingh Pardeshi

Problem Statement

Implement K-means algorithm using MPI. The k-Means algorithm aims to partition N data points into k groups called clusters, such that each data point belongs to a cluster with nearest mean.

Algorithm

1. Generate the data and divide the data equally on to each cluster.

2. Choose initial K data points as the initial centroids.

3. Broadcast the initial centroids to each of the cluster.

4. For each data point calculate the distance of the data point from each centroid to find the nearest cluster.

5. Assign that data point to nearest cluster.

6. Calculate the centroids of the new clusters by averaging data points in the new cluster.

7. Calculate flag = difference between old and new centroids

8. if flag < threshold then go to Step 3. Else stop.

Submission

The submission contains a source code file named **“mpi_k_mean.c”** and the executable file **“mpi_k_mean.exe”** and the output file **“output_12.txt”**