

Напишите две C++ программы, использующие технологию MPI для распределенного решения задачи кластеризации и оценки качества кластеризации.

Подготовьте соответствующий датасет, содержащий численные значения и не менее 1000 записей (например, из датасета Brooklyn sales).

Процесс 0 (root) считывает данные в массив и отправляет их на все процессы в коммутаторе. Остальные процессы запускают алгоритм кластеризации по методу K-средних (K-means clustering) и находят его решение. Когда процесс получает решение, он оценивает его с помощью любого индекса кластеризации (PDF-файл Clustering Quality Measures).

Процесс 0 собирает значения индексов со всех процессов и находит лучшее решение, по критерию минимума или максимума (с.21, Min или Max). Вывести значение центроидов кластеров лучшего решения.

Первая программа (версия программы) использует point-to-point соединение (функции MPI_Send и MPI_Recv), а вторая программа использует функции коллективной коммуникации (например, MPI_Bcast или MPI_Reduce). Сравнить производительность по времени для двух версий программы и сделать выводы.