

Распараллеливания алгоритма кластеризации k-Means

Летняя школа, 2019г.

Нұрберген Қуаныш

Факультет информационных технологии

12.07.2019г.

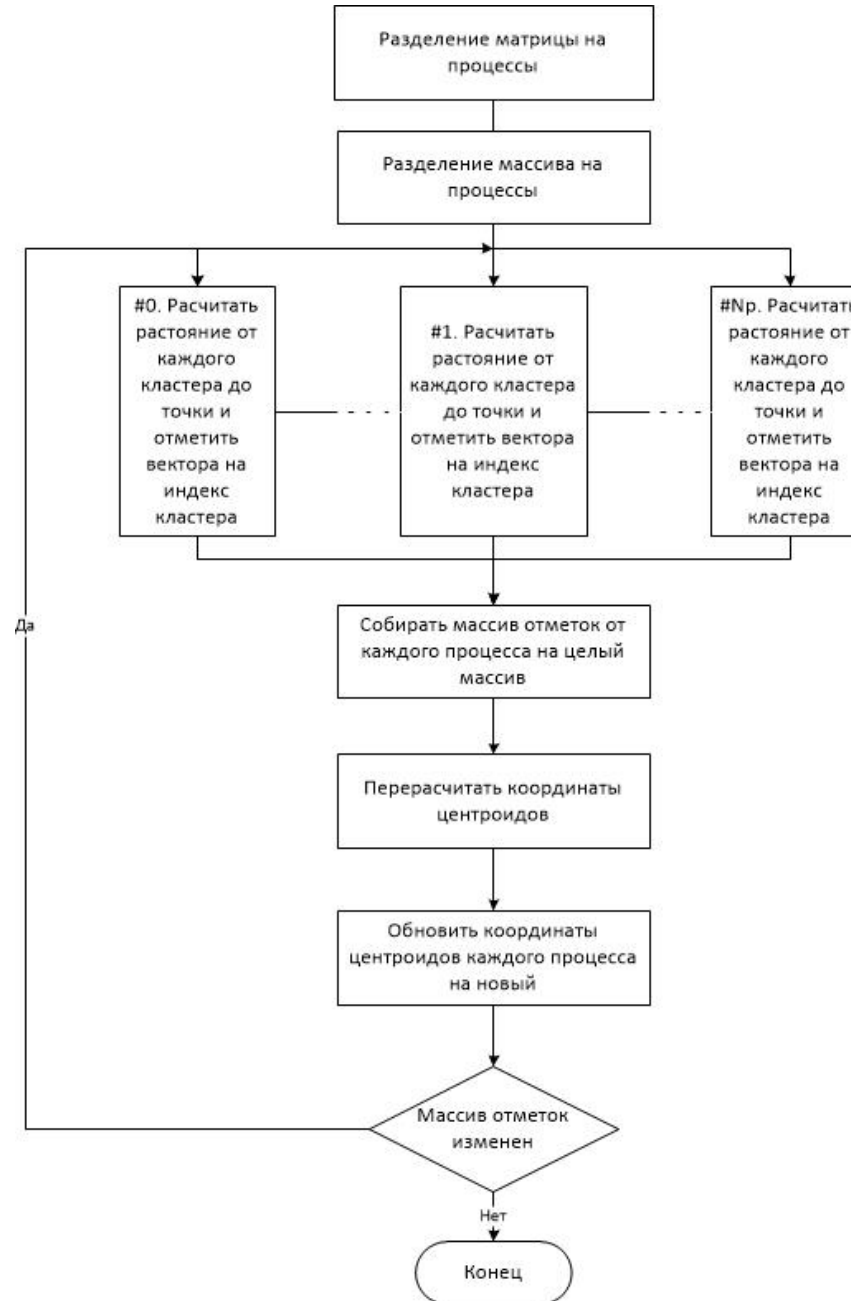
Постановка задачи

- Сделать параллелизацию k-Means алгоритма (алгоритм кластеризации) с MPI технологией
- Реализовывать параллельную программу k-Means на языке C#

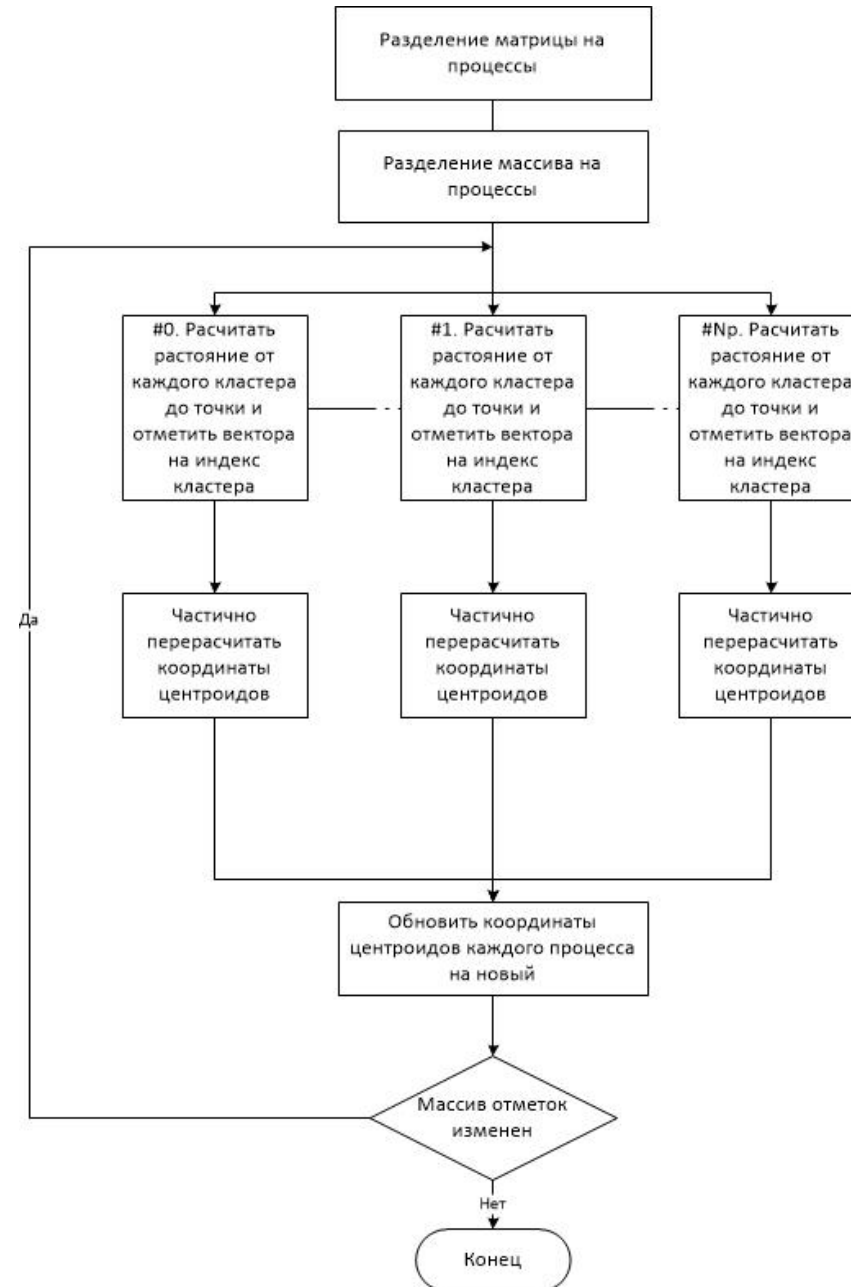
Алгоритм K-Means

- Инициализация начальных координат центроидов
- Объединить точек к одному кластеру, которые близки к центроидам
- Пересчитать координату центроидов
- Проверка. Если ни одна точка не изменил свою кластер то остановить

Решения #1



Решения #2

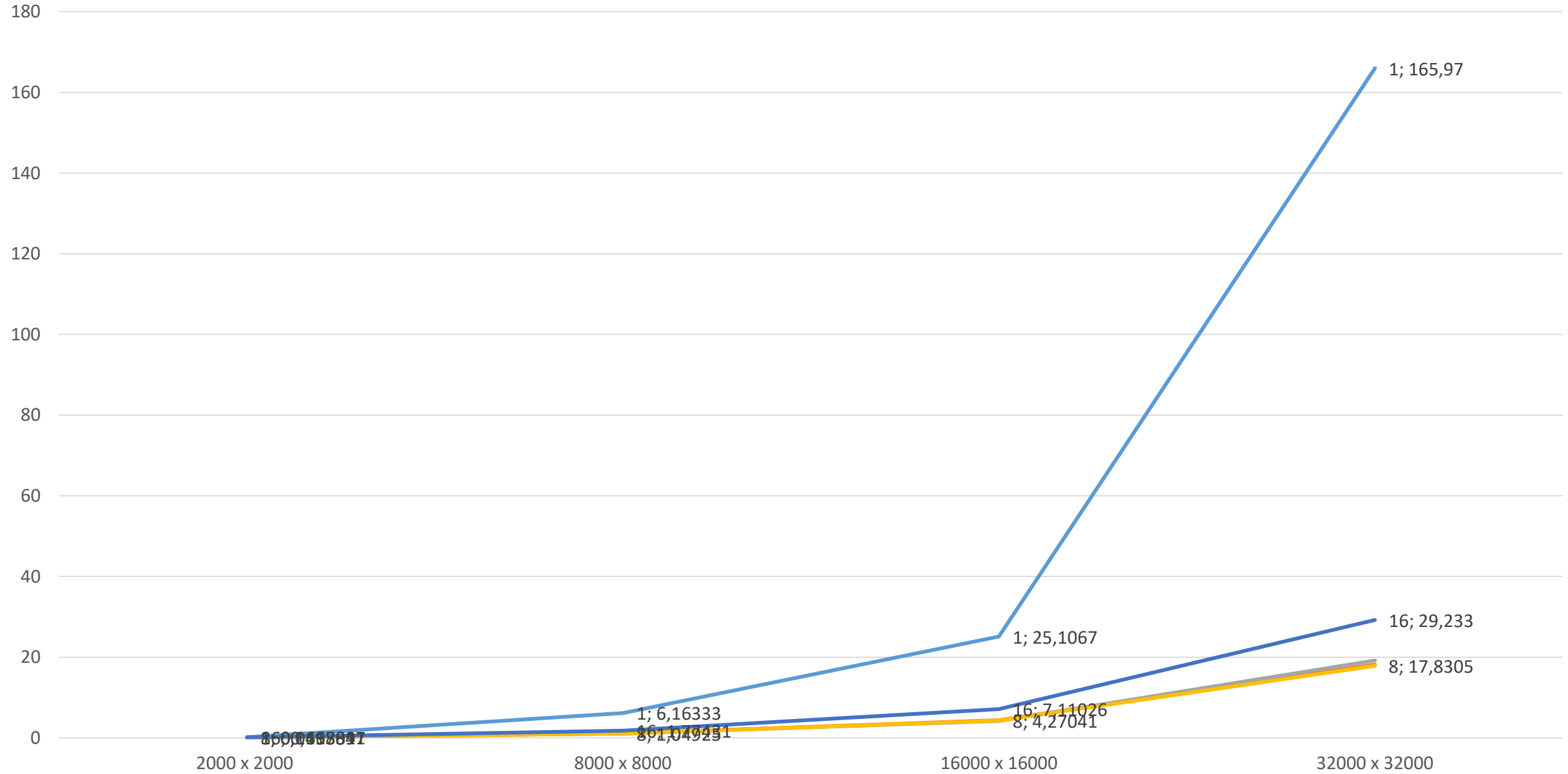


Тестирование MPI программы #1

	Последов.	2 процесс	4 процесс	8 процесс	16 процесс
2000 x 2000	0,146589 сек.	0,0683582 сек.	0,0662701 сек.	0,0698311	0,117047 сек.
8000 x 8000	6,16333 сек.	1,08317 сек.	1,02546 сек.	1,04925	1,77431 сек.
16000 x 16000	25,1067 сек.	4,39636 сек.	4,17156 сек.	4,27041	7,11026 сек.
32000 x 32000	165,97 сек.	18,1407 сек.	19,1517 сек.	17,8305	29,233 сек.

MPI Method 1

1 2 4 8 16

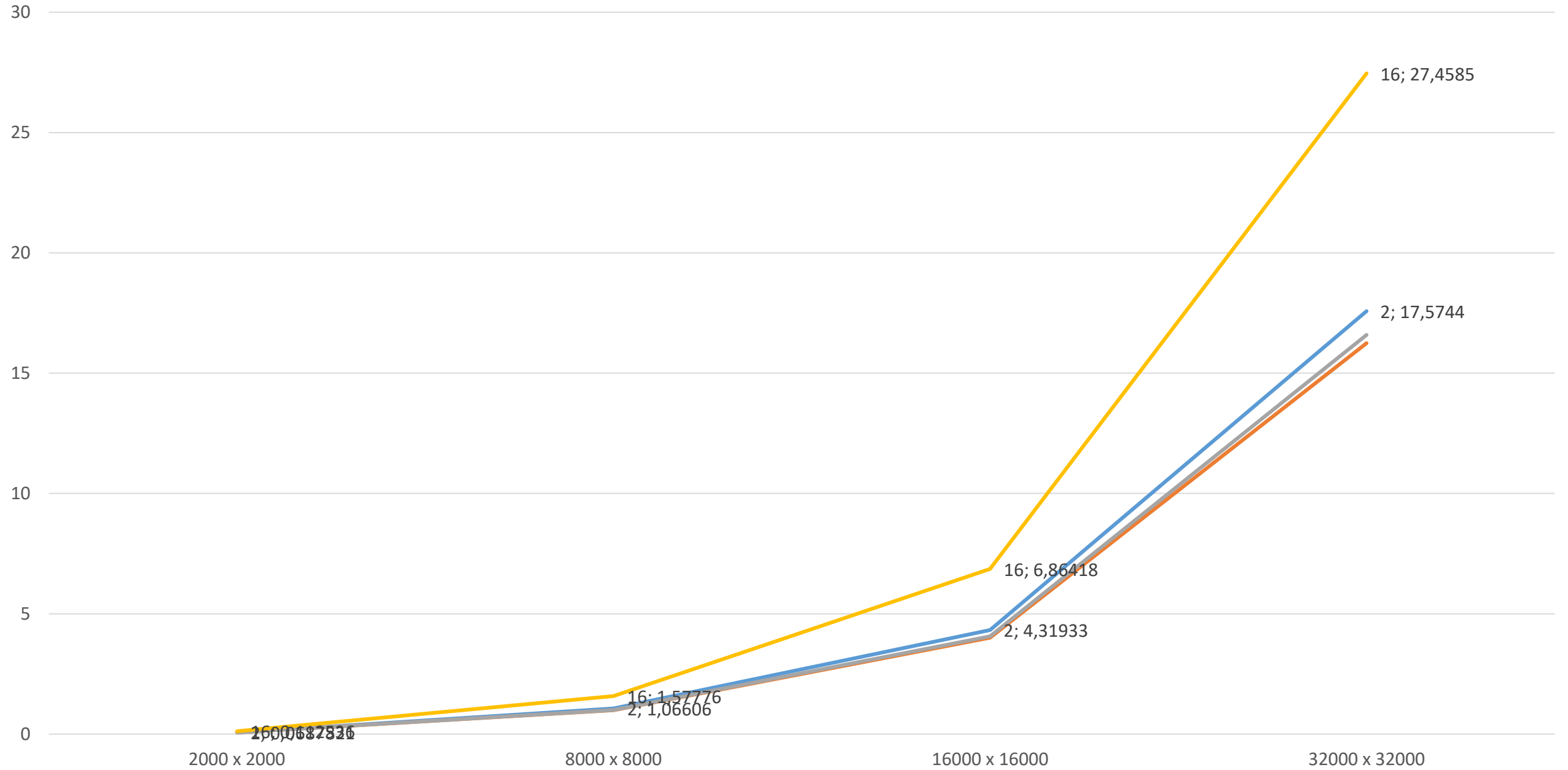


Тестирование MPI программы #2

	2 процесс	4 процесс	8 процесс	16 процесс
2000 x 2000	0,0687821	0,064578	0,067414	0,112536
8000 x 8000	1,06606	0,993937	1,00253	1,57776
16000 x 16000	4,31933	4,00262	4,06453	6,86418
32000 x 32000	17,5744	16,2462	16,5931	27,4585

MPI Method 2

— 2 — 4 — 8 — 16



Заключение

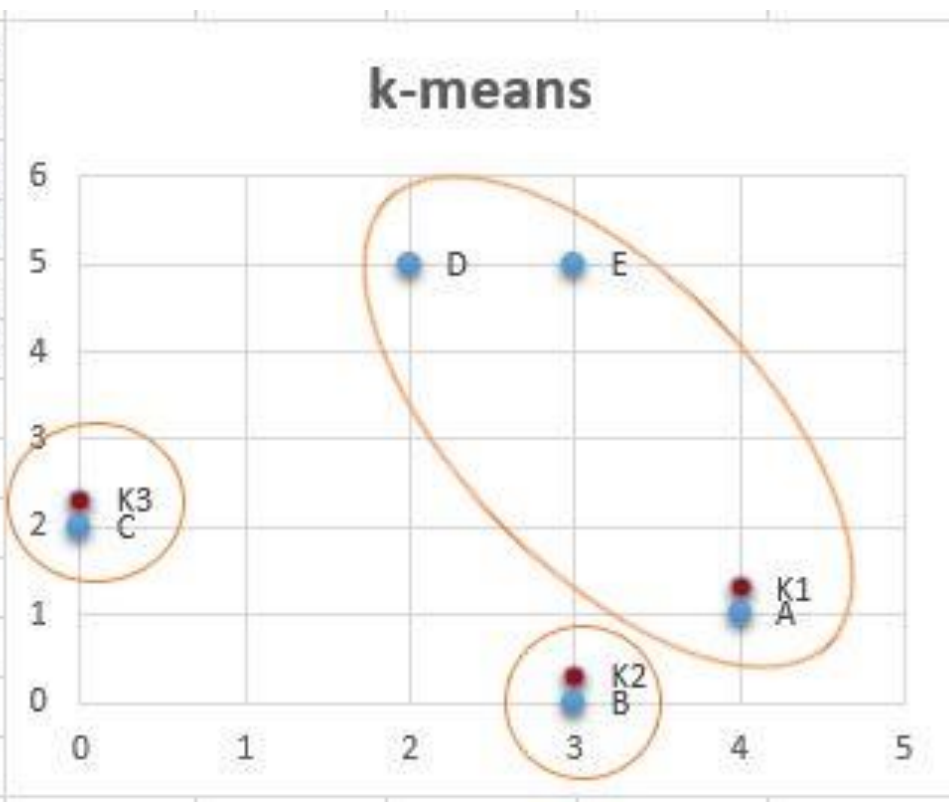
Сделано:

- Сделано параллелизация алгоритма K-Means с MPI технологией.
- Сделано параллелизация алгоритма на C# Task, Thread.

- В будущем планируется улучшить качество кластеризации (машинном обучением, или д.р. методами).

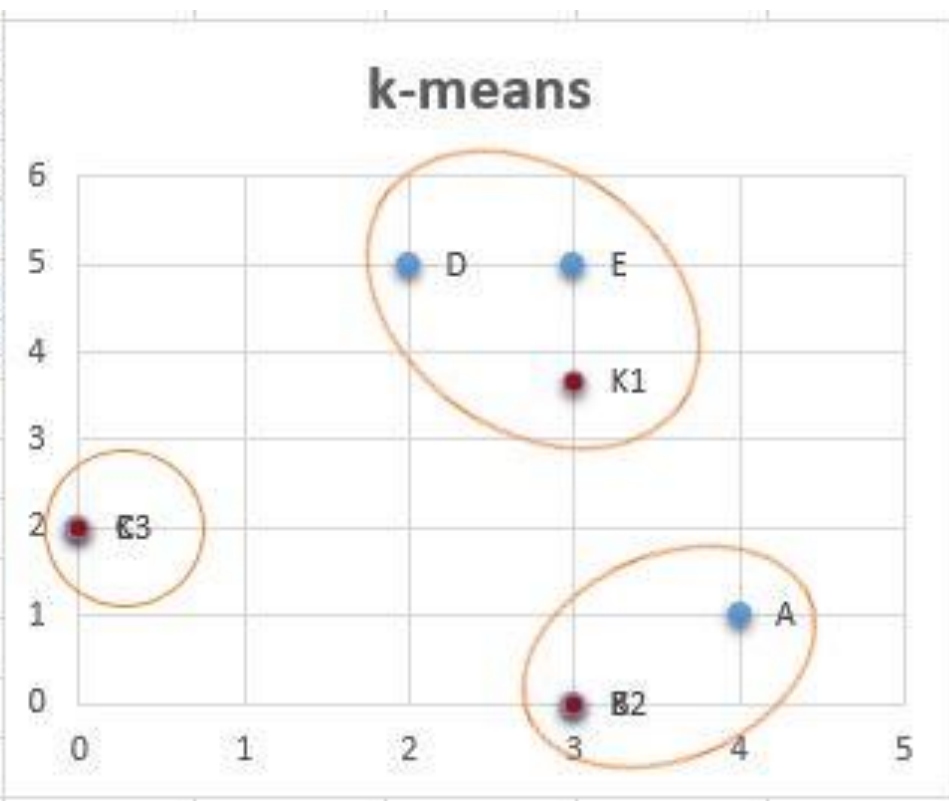
Пример (1-итерация)

Points	x	y	distance K1	distance K2	distance K3	Cluster
A	4	1	0,3	1,220655562	4,20594817	1
B	3	0	1,64012195	0,3	3	2
C	0	2	4,0607881	3,44818793	2	3
D	2	5	4,20594817	4,805205511	5,38516481	1
E	3	5	3,83275358	4,7	5,83095189	1
K1	4	1,3				
K2	3	0,3				
K3	0	2,3				



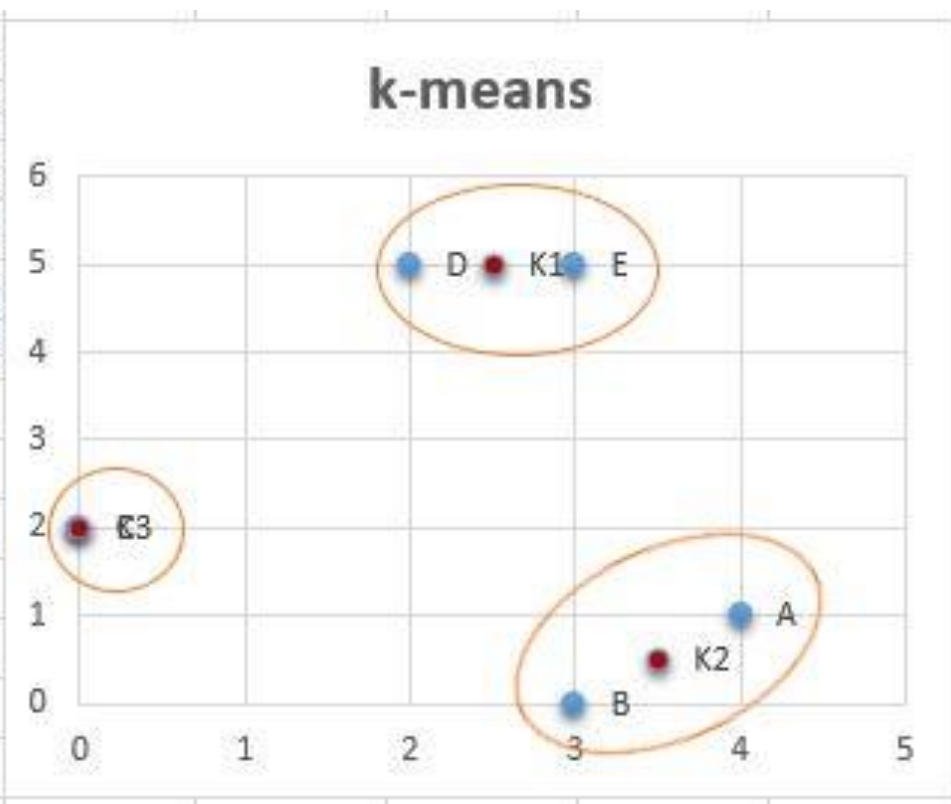
Пример (2-итерация)

Points	x	y	distance K1	distance K2	distance K3	Cluster
A	4	1	2,85112259	1,414213562	4,12310563	2
B	3	0	3,67	0	3	2
C	0	2	3,43349676	3,605551275	2	3
D	2	5	1,6640012	5,099019514	5,38516481	1
E	3	5	1,33	5	5,83095189	1
K1	3	3,67				
K2	3	0				
K3	0	2				

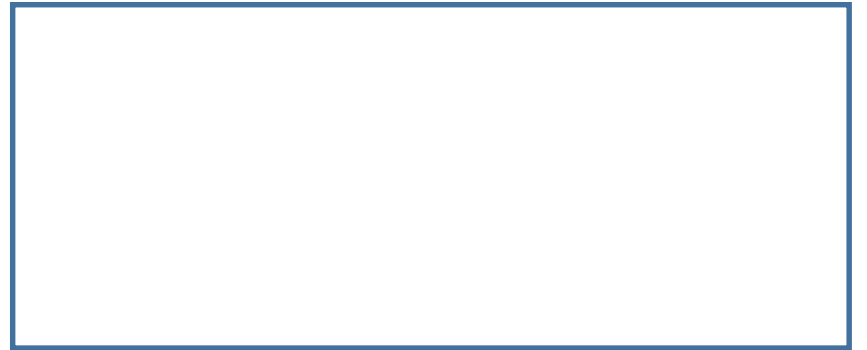


Пример (3-итерация)

Points	x	y	distance K1	distance K2	distance K3	Cluster
A	4	1	4,27200187	0,707106781	4,12310563	2
B	3	0	5,02493781	0,707106781	3	2
C	0	2	3,90512484	3,807886553	2	3
D	2	5	0,5	4,74341649	5,38516481	1
E	3	5	0,5	4,527692569	5,83095189	1
K1	2,5	5				
K2	3,5	0,5				
K3	0	2				



Решения #1



Решения #2

