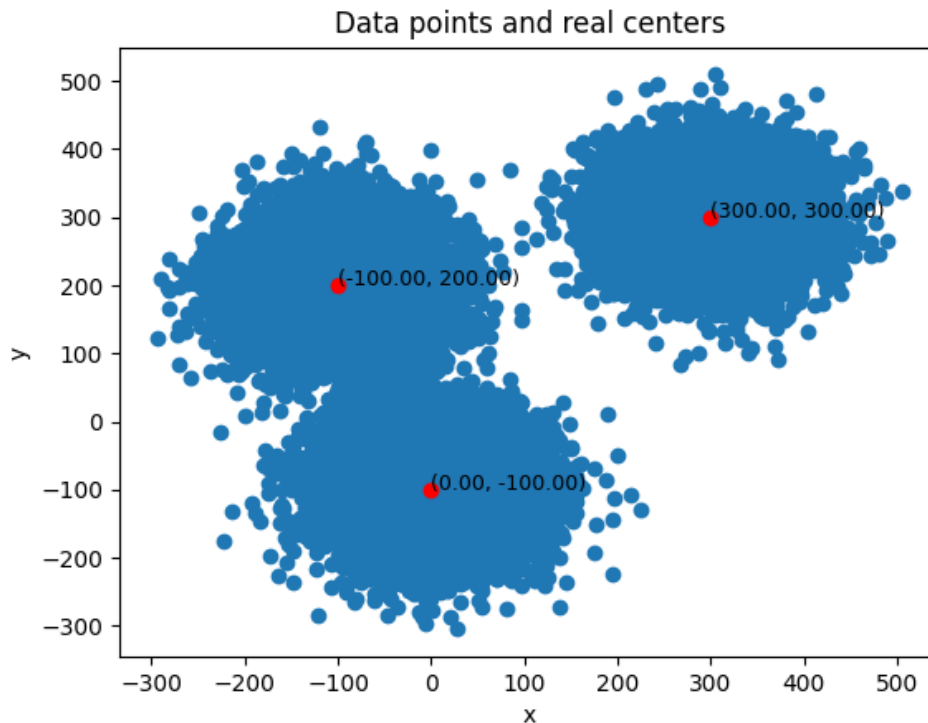


Hadoop Map-Reduce Project

KMeans algorithm implementation

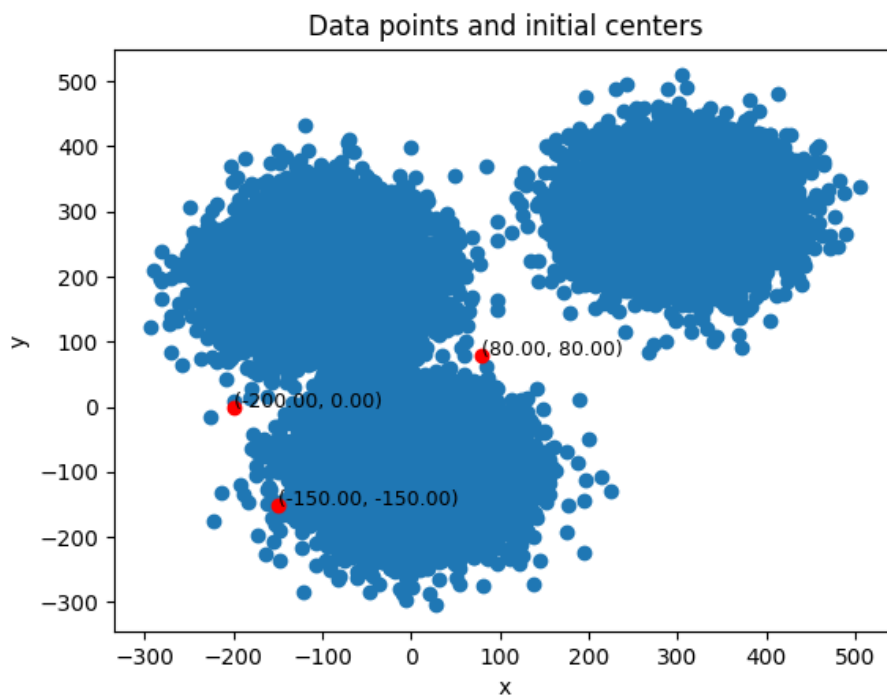
Christos Patrinopoulos p3200150

Visualization of data points created by Python script *createPoints.py* initially.
Centers used: (-100, 200), (300, 300), (0, -100)



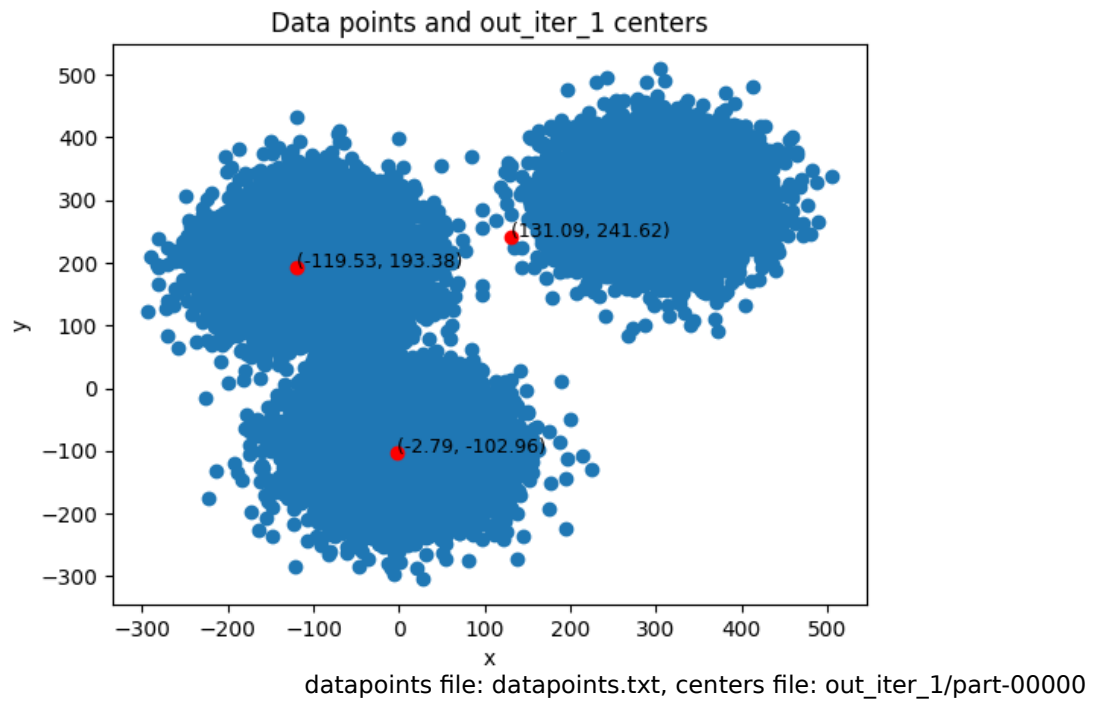
datapoints file: datapoints.txt, centers file: real-centers.txt

The centers used as an input to the system were: (-200, 0), (80, 80), (-150, -150)
The KMeans iteration ran 4 times. The procedure stops if the distance difference between the old and new calculated centers is less than 0.01.

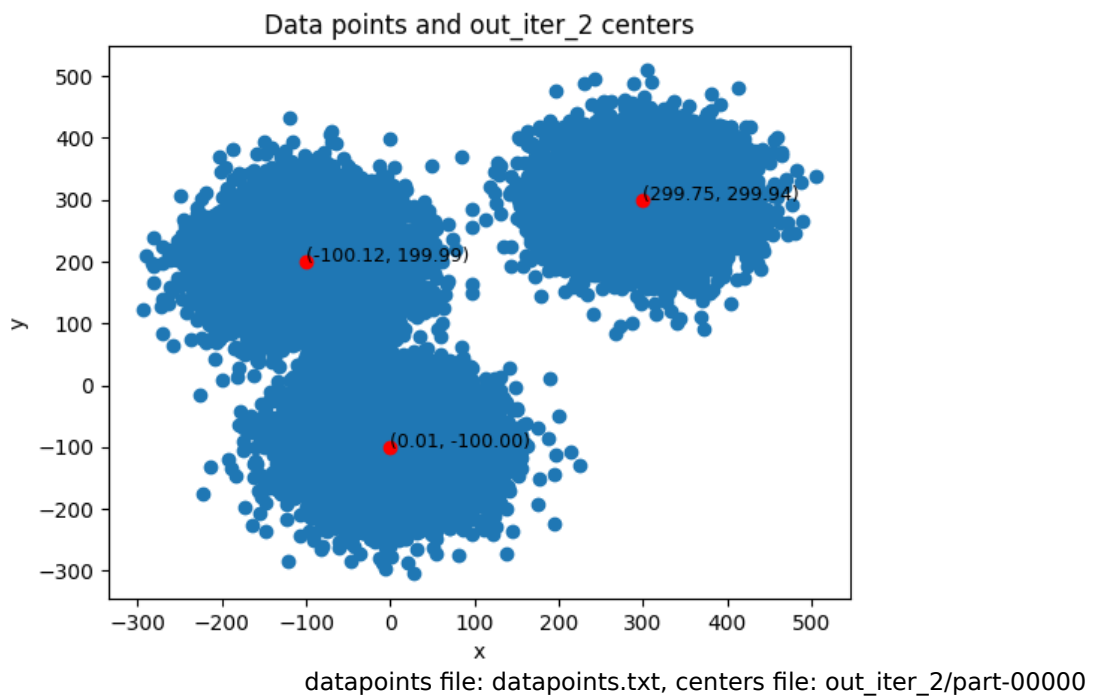


datapoints file: datapoints.txt, centers file: init-centers.txt

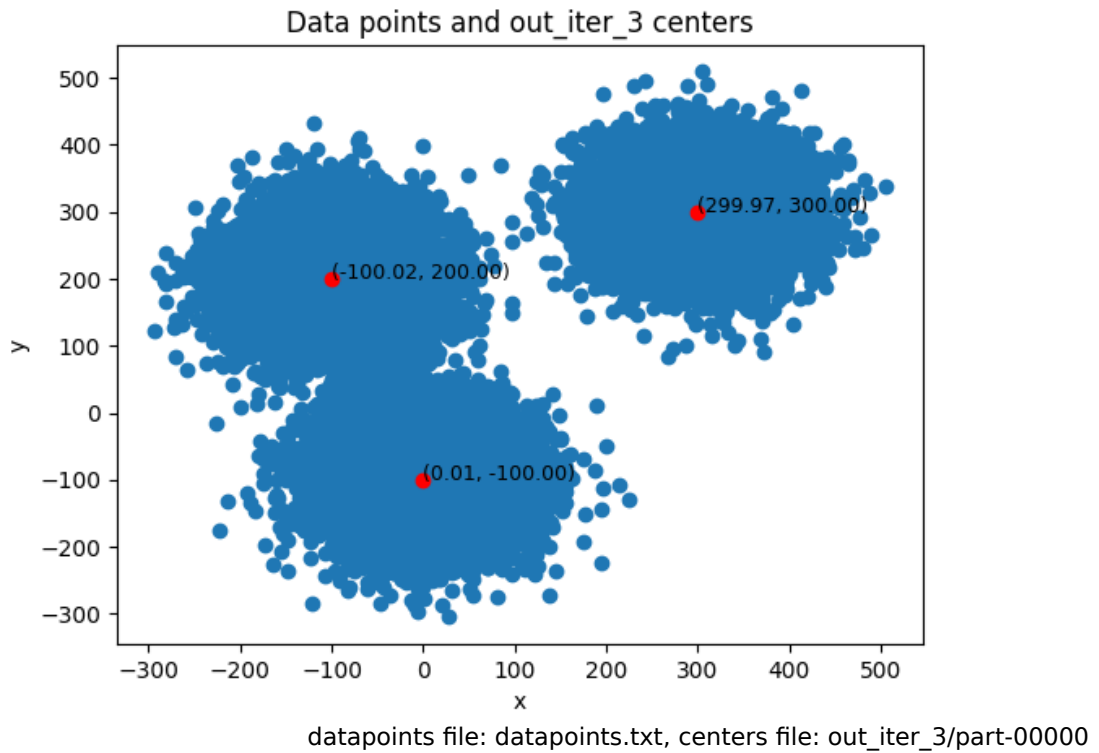
After iteration 1:



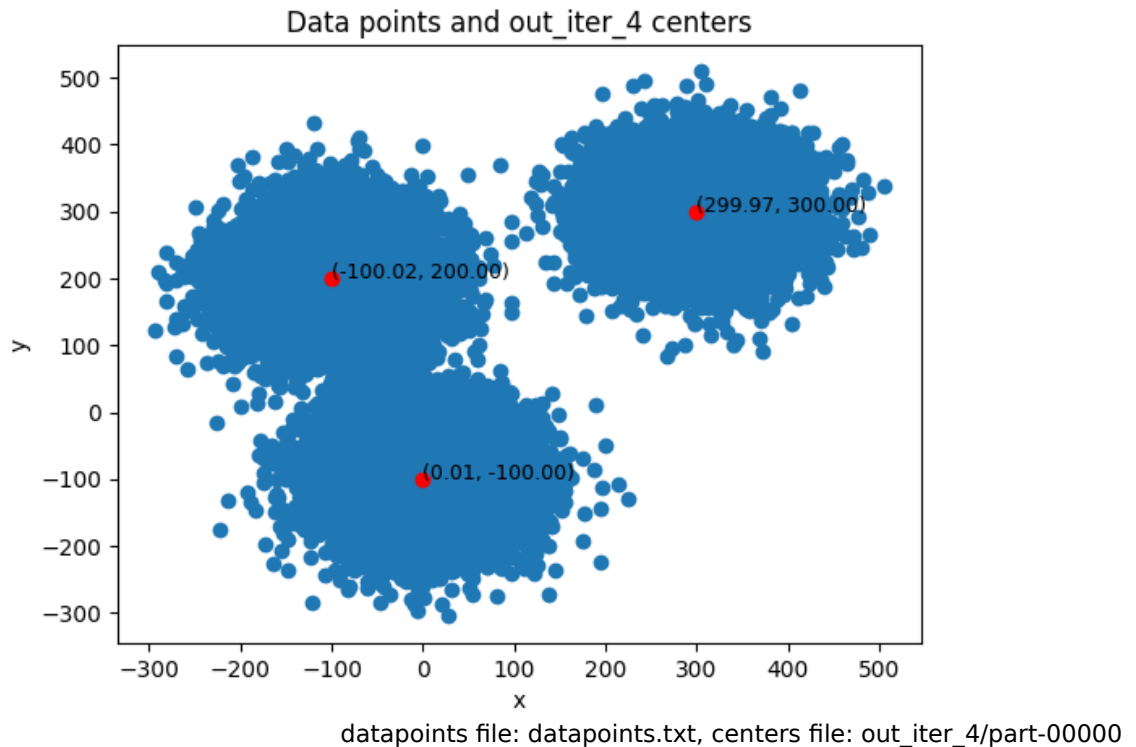
After iteration 2:



After iteration 3:



After iteration 4:



As we can see after the 4th iteration of the KMeans algorithm in Hadoop, the centers are almost identical to the real centers we used to create the data points. KMeans stops running.

*Input folder: datapoints.txt, real-centers.txt, init-centers.txt
Outputs can be found in directory: output-2024-04-16-19-55-56/output*