# Group 13 Homework 1 Analysis

Marco Annunziata, Sveva Turola

We denote as:

* N: size of the input.
* L: the number of partitions;
* , meaning: the maximum size across all partitions;

Design Goals for the MapReduce algorithm implemented:

* R (number of rounds): 2
* ML (local space):
  + Round 1:
    - Map Phase: so that each element is transformed into a pair , where is the distance between point and the closest center;
    - Reduce Phase: , since the pairs are divided into L partitions of size *m* and the partial counts of the points with the same closest center are computed for each partition, meaning that the amount of space needed for the intermediate pairs is *m*.
  + Round 2:
    - Map Phase: *empty*;
    - Reduce Phase: , where for each center the partial counts of all the partitions are summed up, so the allocated space increases linearly with the number of partitions.

Overall Local Space:

If , then by using random partitioning there is (high) probability that:

And so,

* MA (Aggregate Space): , since in every round the input/intermediate/output pairs require space overall.