

Homework 4

Department of Information Engineering
Big Data Computing
Prof. Andrea Pietracaprina
Due Date: 31/05/2018

Group 18

Aladdine Ayadi
Giovanni Barbieri
Alessandro Pelizzo
Davide Talon

Test results

The present report shows the obtained results for the 4 – *th* homework which has the goal to study the performances of a cluster in the Diversity Maximization problems. In particular the our experiment focus on the

Let N_{exec} be the number of cores per executor for a total number of N_{tot} cores used. The measurements evaluate the time t_c which denotes the time for the construction of the coreset, t_s represents the time for the run of sequential algorithm and finally t_l the time needed for the loading and counting of the dataset, all times are expressed in ms.

Table 1: Results with dataset *all*, $P = 32$, $k = 20$ and increasing cores.

N_{tot}	N_{exec}	$t_c[ms]$	$t_s[ms]$	$t_l[ms]$
4	2	14785	165	169256
8	2	7262	194	92361
8	4	12393	174	114576
8	8	23183	324	229915
16	4	22651	229	147105
16	8	22902	212	215455
32	4	3037	195	33901
32	8	24426	516	227701
64	4	19192	299	148242
64	8	22688	188	206375

Table 2: Results with dataset *all*, $N_{tot} = 32$, $N_{exec} = 4$, $k = 20$ and increasing P .

P	$t_c[ms]$	$t_s[ms]$	$t_l[ms]$
4	12402	21	346732
8	7808	36	42784
16	4885	80	40415
32	3037	195	33901
64	9715	607	44835
128	4139	2050	43149
256	5222	8189	38288

Table 3: Results with $N_{tot} = 32$, $N_{exec} = 4$, $k = 20$, constant number of partitions P and increasing size of the dataset.

Dataset	$t_c[ms]$	$t_s[ms]$	$t_l[ms]$
vectors-50-500000	1848	170	26885
vectors-50-1000000	2190	193	31254
vectors-50-2000000	5325	179	73585
vectors-50-3000000	2486	232	40012
vectors-50-all	3037	195	33901