**BIG DATA COMPUTING 2018-19 – HOMEWORK 4 – GROUP 11**

1. **Required tests.** Do the tests with the parameters indicated in the following table and, for every test, report the following values: Ti = running time (in seconds) of Round i of MR\_kmedian (i=1,2,3), and Obj = value of the objective function (average distance of points from centers). If you notice anomalies in the values of a test try to repeat the test 2-3 times and take the average values*. Some anomalies are to be expected!*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **K** | **L** | **iter** | **num-executors** | **T1** | **T2** | **T3** | **Obj** |
| **HIGGS11M7D** | **10** | **16** | **0** | **16** | **7.132** | **0.599** | **0.893** | **0.615779066** |
| **HIGGS11M7D** | **25** | **16** | **0** | **16** | **14.512** | **0.312** | **3.878** | **0.4896777433** |
| **HIGGS11M7D** | **50** | **16** | **0** | **16** | **36.689** | **0.714** | **1.723** | **0.422128416** |
| **HIGGS11M7D** | **50** | **16** | **10** | **16** | **50.737** | **0.537** | **3.586** | **0.360834049** |
| **HIGGS22M7D** | **50** | **32** | **10** | **8** | **81.496** | **0.651** | **7.7** | **0.432555544** |
| **HIGGS22M7D** | **50** | **32** | **10** | **16** | **62.21** | **0.614** | **7.485** | **0.432571833** |
| **HIGGS22M7D** | **50** | **32** | **10** | **32** | **57.285** | **1.455** | **6.334** | **0.432463458** |

1. **Other observations** (at your discretion)

We used cache() on the first round, this choice led us to have a fast second round (because collect() is lighter) and a slower first round.