**BIG DATA COMPUTING 2019-20 – HOMEWORK 3 – GROUP 28**

**(USE ONLY 1 PAGE)**

1. **Required tests.** Do the tests with the parameters indicated in the first 4 columns of the following table. Fill the table reporting, for each test, the following values: **Init** = *time to read the input and create the RDD*;**Ti** = *time of Round i of runMapReduce* (i=1,2), **AvgDist** = *average distance among the solution points*. **Times should be in ms.** If you notice anomalies in the values of a test try to repeat the test 3 times and take the median values*.*

*Some anomalies are to be expected!*

**Table to be filled by Java users**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **K** | **L** | **num-executors** | **Init** | **T1** | **T2** | **AvgDist** |
| Glove2M300d.txt | 100 | 16 | 16 | 24034 | 7 | 32251 | 29.040733847821866 |
| **Glove2M300d.txt** | **100** | **16** | **8** | 34498 | 7 | 40466 | 29.040733847821866 |
| **Glove2M300d.txt** | **100** | **16** | **4** | 68536 | 11 | 66170 | 29.040733847821866 |
| **Glove2M300d.txt** | **100** | **4** | **4** | *You will not be able to run this test* | | | |

**Table to be filled by Python users**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **K** | **L** | **num-executors** | **Init** | **T1** | **T2** | **AvgDist** |
| **Glove200K300d.txt** | **30** | **16** | **16** |  |  |  |  |
| **Glove200K300d.txt** | **30** | **16** | **8** |  |  |  |  |
| **Glove200K300d.txt** | **30** | **16** | **4** |  |  |  |  |
| **Glove200K300d.txt** | **40** | **16** | **16** |  |  |  |  |

1. **Observations:** *write here a brief comparison among the results of the experiments reported in the table (and perhaps of other experiments that you did on your own), and try to justify the differences that you may have observed.*