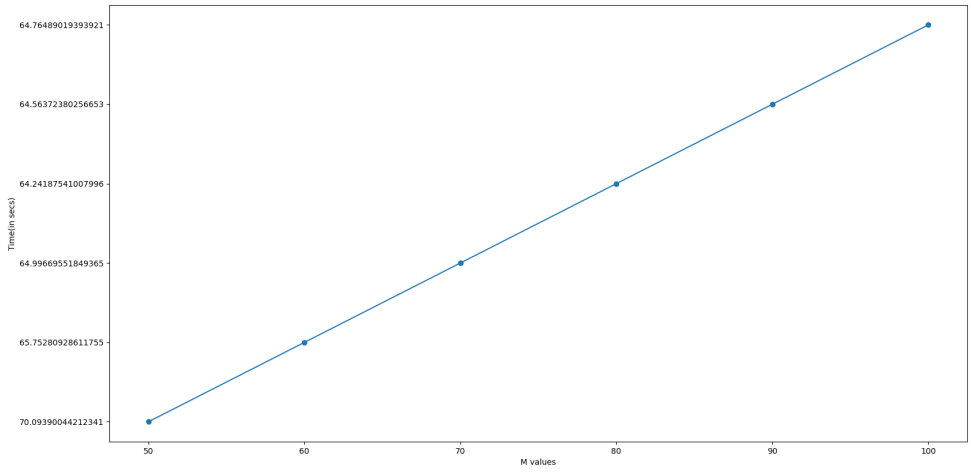
1 Configuration

- Windows 10
- Intel Core i5-4570TE Processor 2 x 2.7 GHz
- 4GB RAM
- Python 3.6

2 Observations

- Size of relation R - 10000 tuples
- Size of relation S - 20000 tuples
- Sort-merge Observations

M	Time(in seconds)
50	70.09
60	65.75
70	64.99
80	64.24
90	64.56
100	64.76

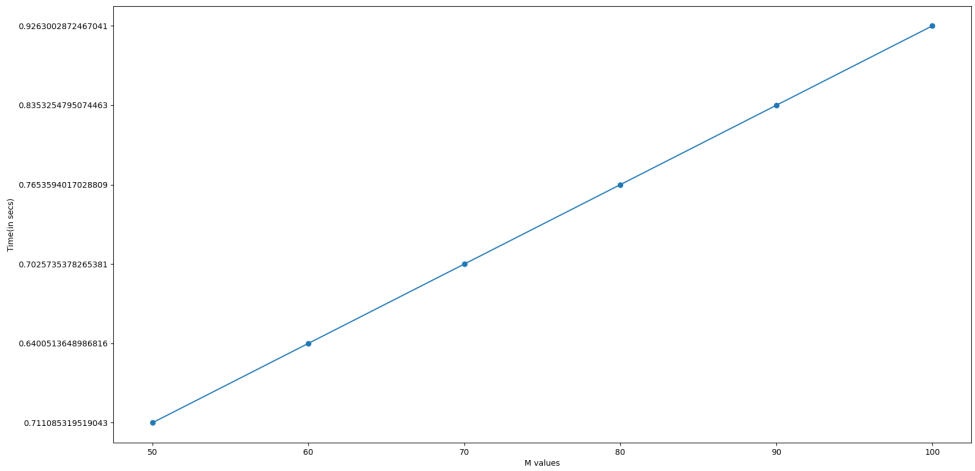


The time taken to join the relations remains roughly same because it is single threaded. First both relations are external sorted. During this each

tuple is read twice. Then during joining each tuple is read only once and this does not change for any M.

- Hash-join Observations

M	Time(in seconds)
50	0.711
60	0.640
70	0.702
80	0.765
90	0.835
100	0.926



The time taken to sort the file increases only slightly for different M values. It is significantly faster than sortmerge. Both R and S are divided into M/2 buckets each. In each bucket there are blocks that can store M/2*100 tuples. If one block fills up, then a new one is created. A loop is created over buckets where for each block of the smaller relation all blocks belonging to corresponding bucket of bigger relation is traversed and using dictionary, join is performed.