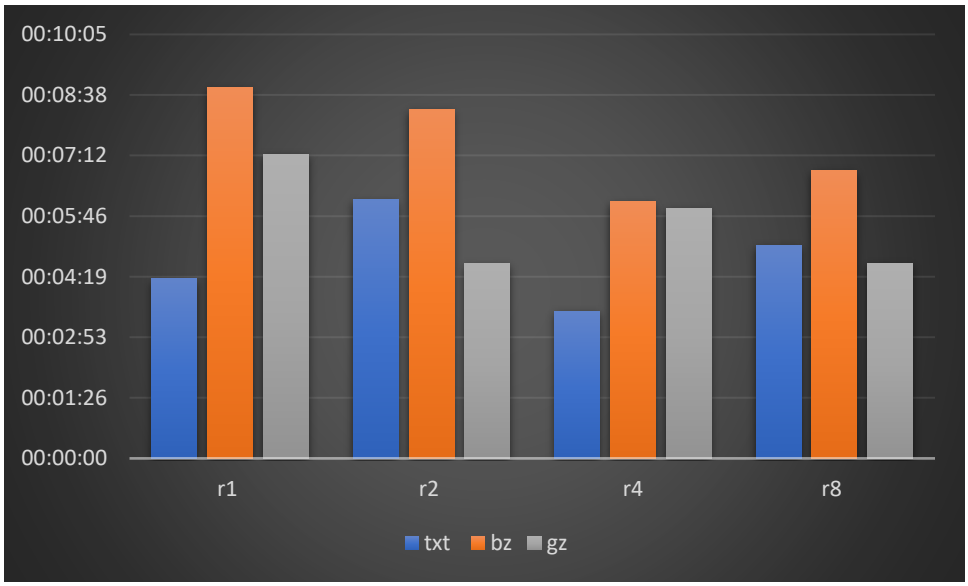
Test-1:



With the Reducers:

The time overall time decreased with increase in number of reducers.

Increase in reducers would decrease the network congestion. Hence the time decreased.

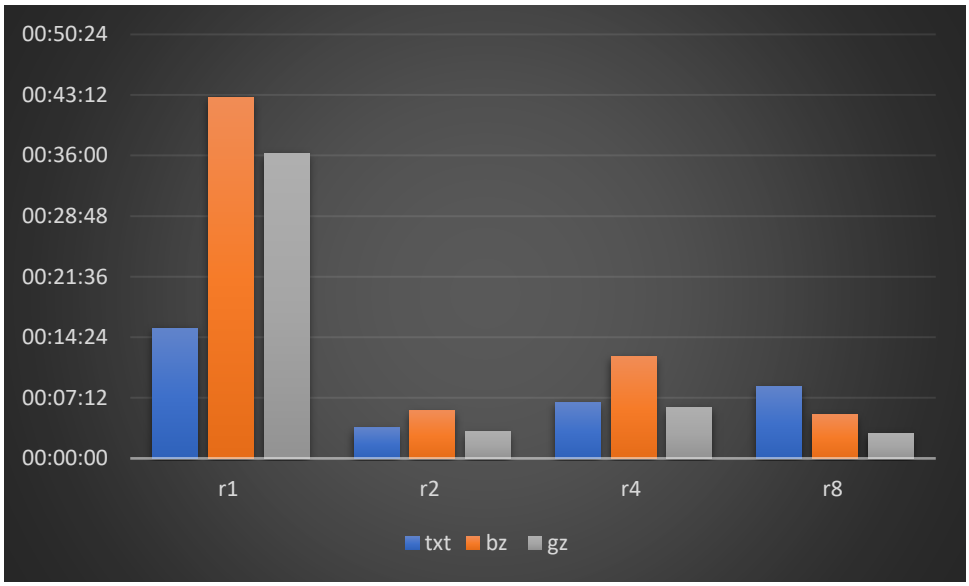
Without Compression:

Since the MapReduce is ran without compression. The time taken to process the compressed bz2 and .gz files is more than text file.

Without Combiner:

No effect on overall output

Test-2:



With Reducers:

The program with least reducers took more time than 2, 4 and 8 reducers.

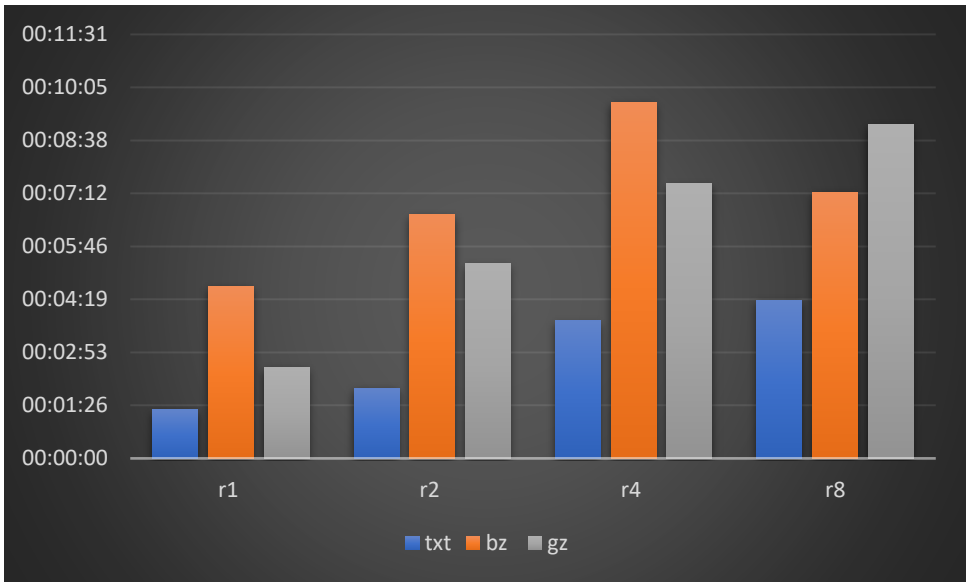
With Compression:

With intermediate compression the time of execution for both the compressed files reduced drastically. The overall time too reduced.

With Combiner:

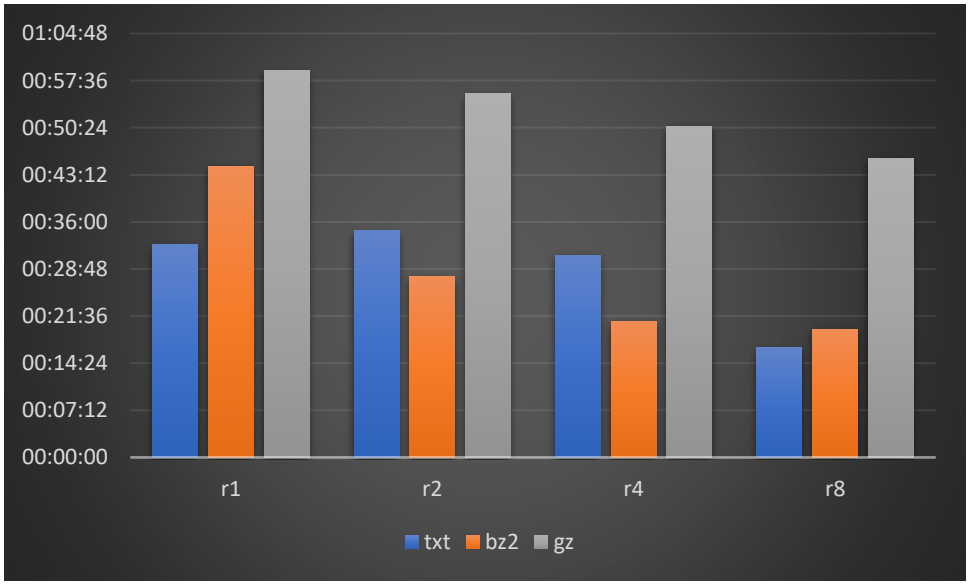
The Reducers execution time reduced because of Combiner

Test-3:



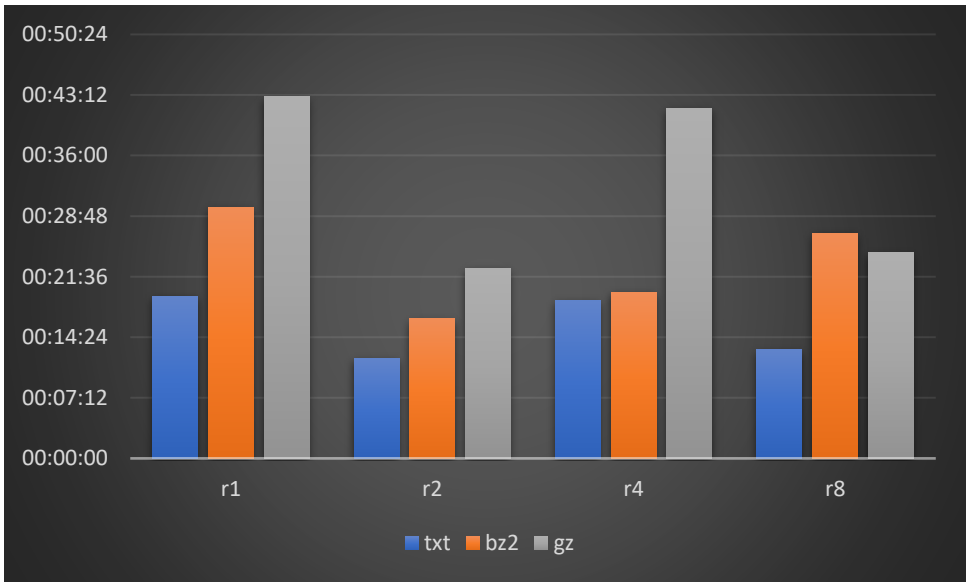
Without Compression, With Combiner, With Reducer the times for the bz2 and gz compressed files increased.

Test – 4:



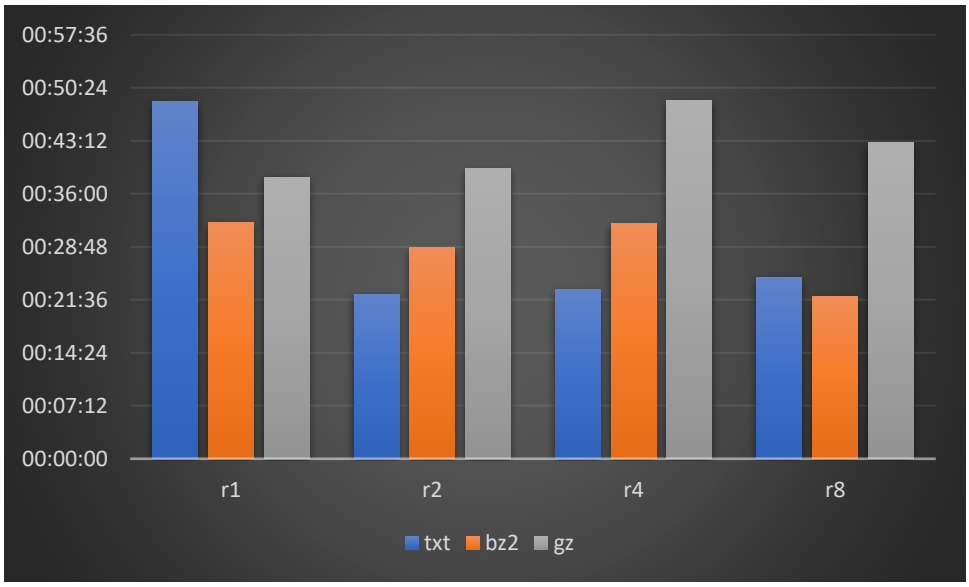
The bad records program without compression and without combiner in test 4 gave an optimized performance with the txt and then with the bz2 file. We also see that the performance was better for a txt file with 8 Reducers, and there was an impact on the performance of the job with the gz file when run with 1 reducer.

Test – 5:



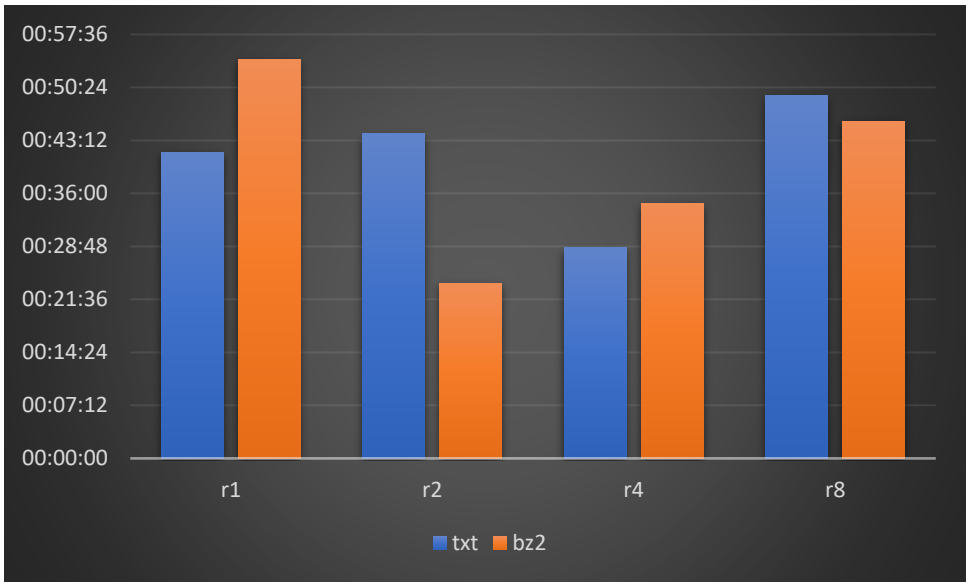
The bad records program with compression and with combiner in test 5 gave an optimized performance with the txt and then with the bz2 file. We also see that the performance was better for a txt file with 8 Reducers and 2 Reducers, and there was an impact on the performance of the job with the gz file when run with 1 reducer.

Test – 6

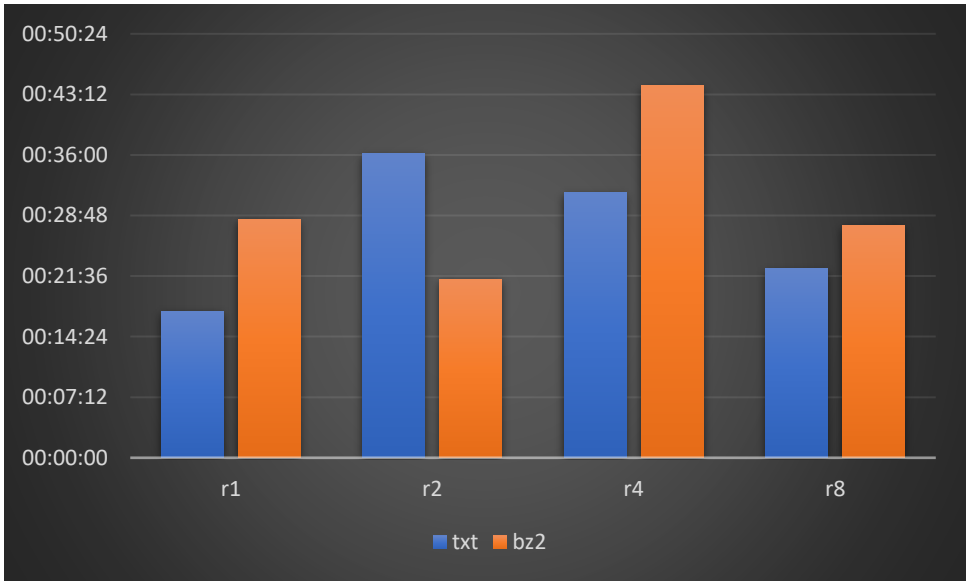


The bad records program with compression and with combiner in test 5 gave an optimized performance with the bz2 file. We also see that the performance was better for a bz2 file with 2 Reducers and 4 Reducers, and there was an impact on the performance of the job with the txt file when run with 1 reducer.

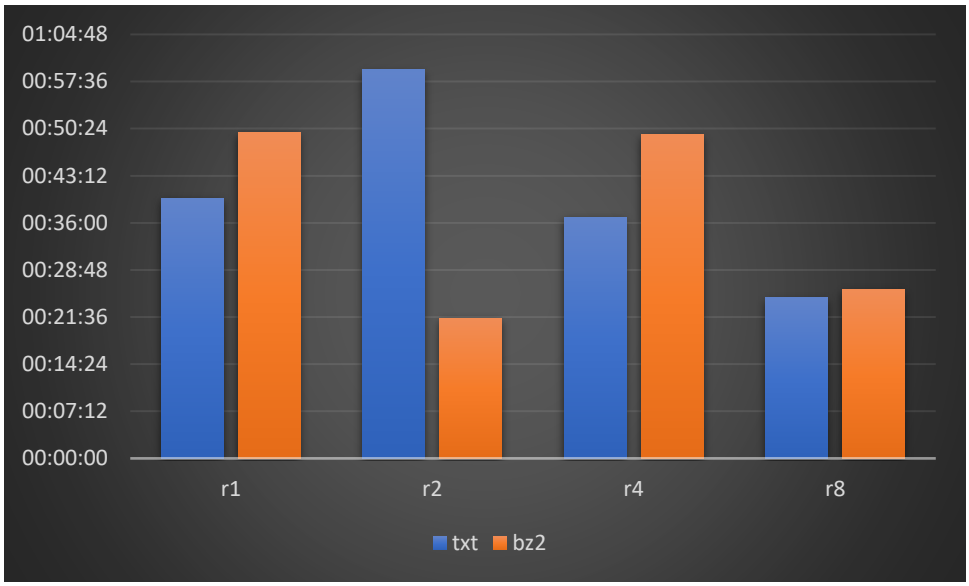
Test – 7:


00:57:36
00:50:24
00:43:12
00:36:00
00:28:48
00:21:36
00:14:24
00:07:12
00:00:00
r1
r2
r4
r8
txt
bz2

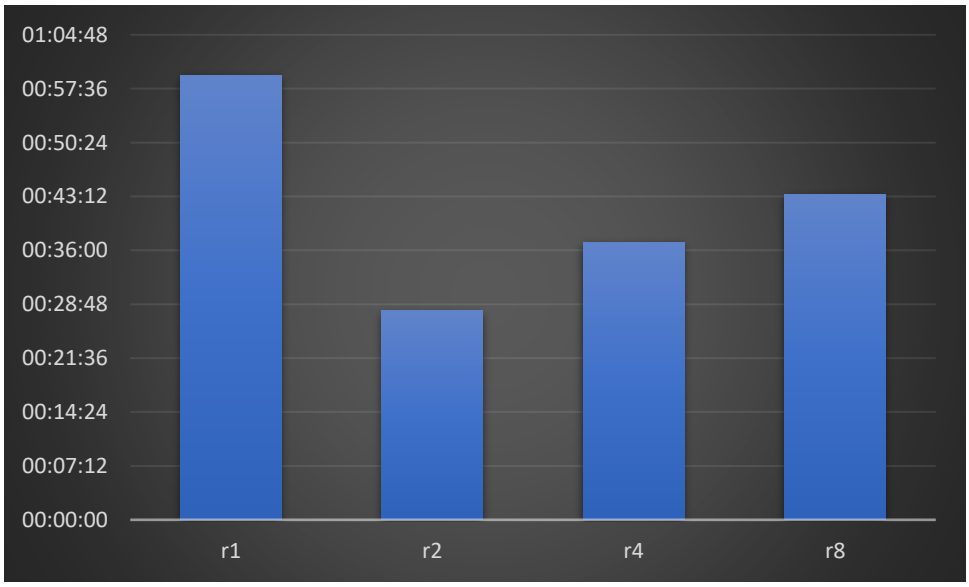
Test-8


00:50:24
00:43:12
00:36:00
00:28:48
00:21:36
00:14:24
00:07:12
00:00:00
r1
r2
r4
r8
txt
bz2

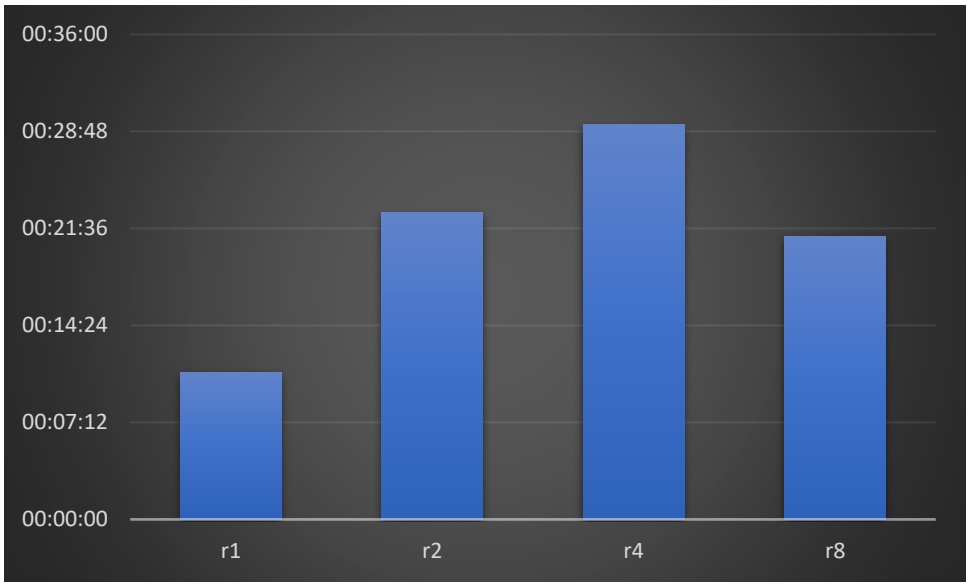
Test-9



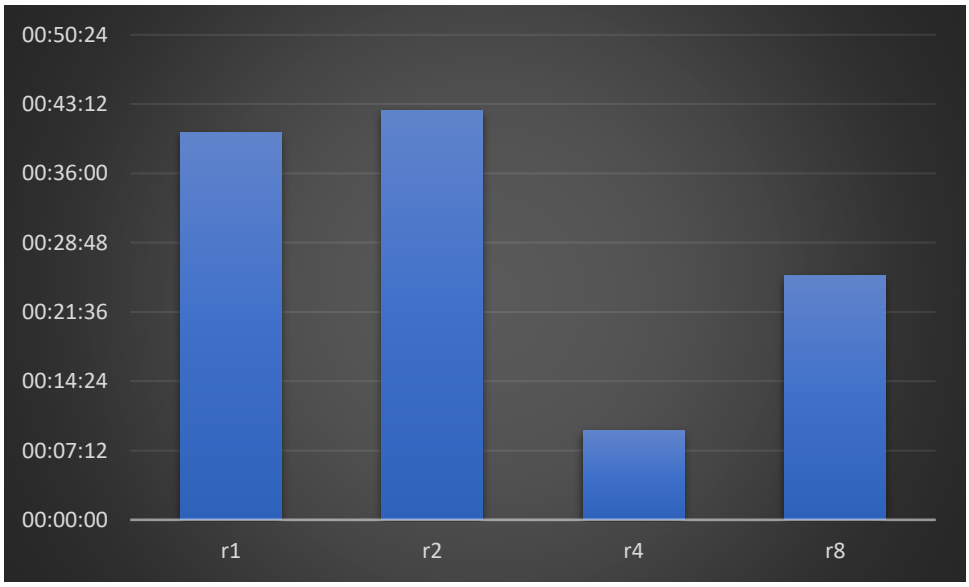
Test-10



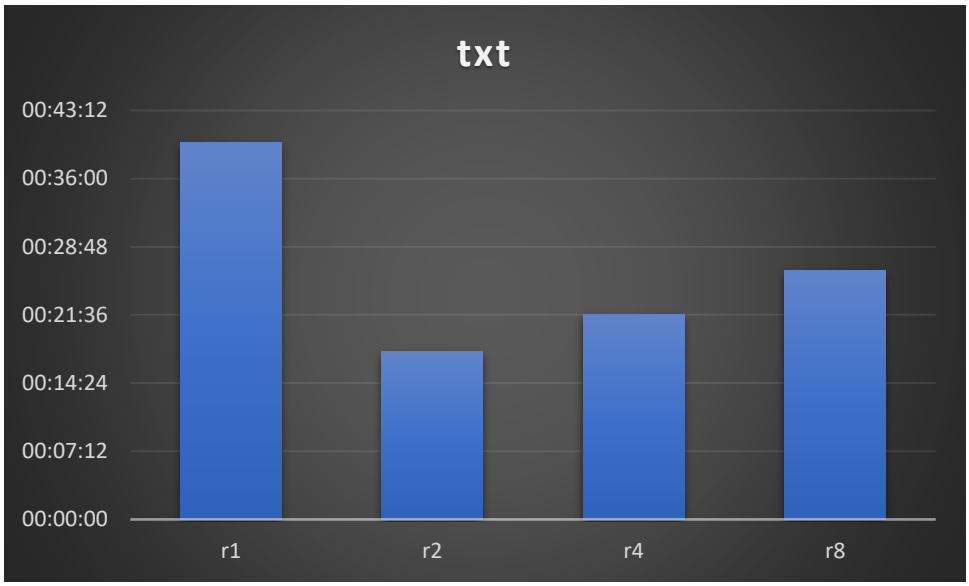
Test -11



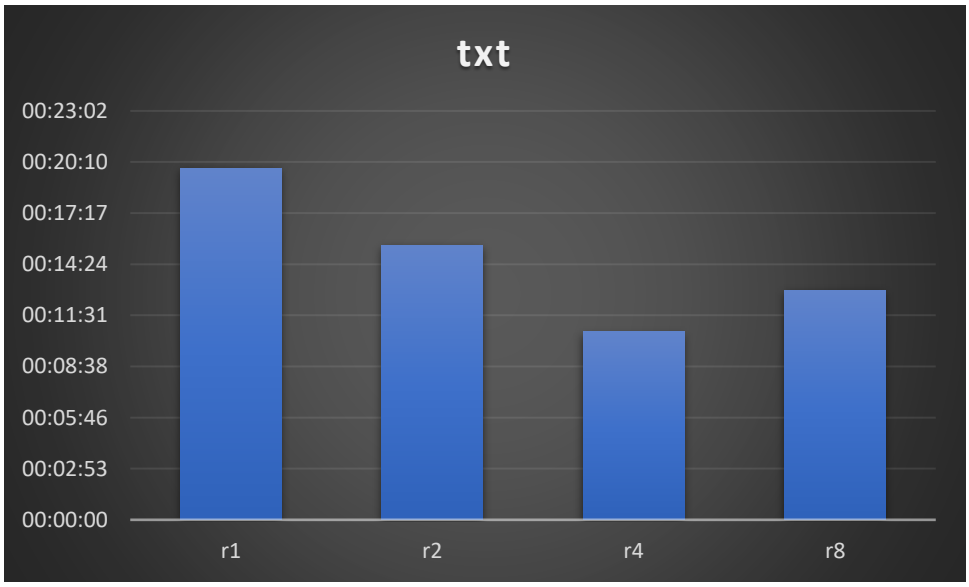
Test -12



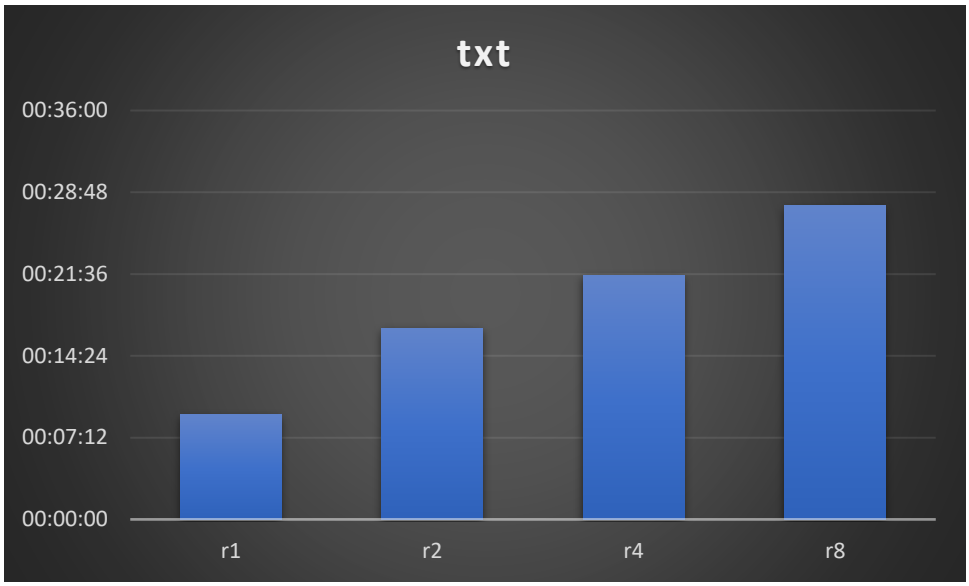
Test – 13



Test – 14



Test – 15



Test – 16 : should run the files with Combiner and Intermediate compression and 8 reducers

Test – 17:

Conclusion:

Number of mappers should be fewer and longer running. Small mappers will be an overhead. We should always prefer multiple reducers to decrease the network congestion. In additional to choosing multiple reducers when we add a combiner then the load on reducer is further decreased. Intermediate Compression we compress data at a mapper output and the volume of data transferred to Reducer is reduced. With the increase in the block size we can reduce the meta data storage and MapReduce prefers large chunk of files rather than small files. This is the reason in our tests small files are taking forever to complete.

Hence, I suggest with combiner, with compression and 8 reducers is the best for smaller Execution times.

Bibliography:

Hadoop Definitive Guide – page 226,175, 217, 34 ,108, 123, 201