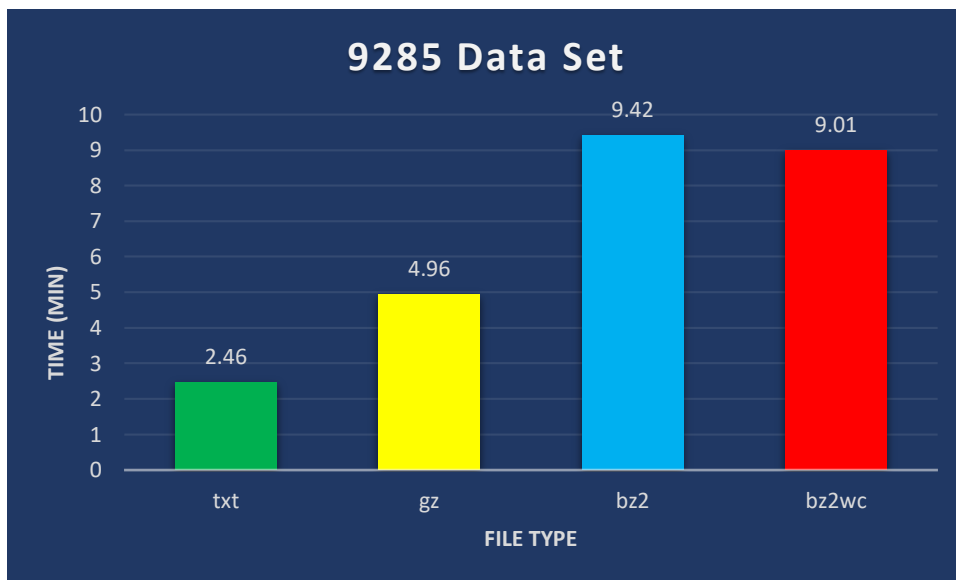
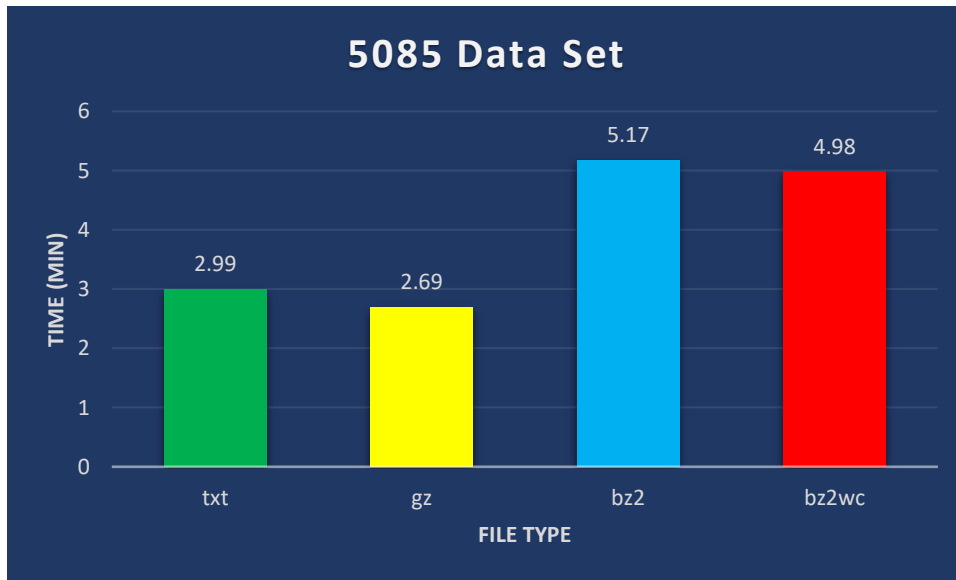


**Compression:**



**Ans:** The similarities between the result is that the gz file format takes the lesser time to get done with the job when compared to bzip2 format. The difference is the time difference as the bzip2 process does splitting whereas, no splitting is performed by the gz compression format. Even increasing the number of reducers only makes the reduce phase go shorter, since we get more of parallelism. However, if we take this too far, we can have end up with lots of small files, which is suboptimal.

Now modifying the bzip2 code by adding combiner in it helps us in reducing some more runtime of job as adding combiner is similar to the addition of a more reducer so basically that's the reason our runtime is getting reduced further.