***SHASHIDHAR NANDADAPU***

***NAU ID-5985492***

**HOMEWORK ASSIGNMNENT-1(LSDS)**

**A.**Using the 1 million reads , the time and RAM it took to initialize 1 million reads is shown below .

That is 6 seconds of CPU usage and 0.0M of RAM

NOTE: we are getting 0 RAM usage for 1million because slurm is polling the program on 30second interval

So,by estimating the time and RAM usage for 36Million we get,

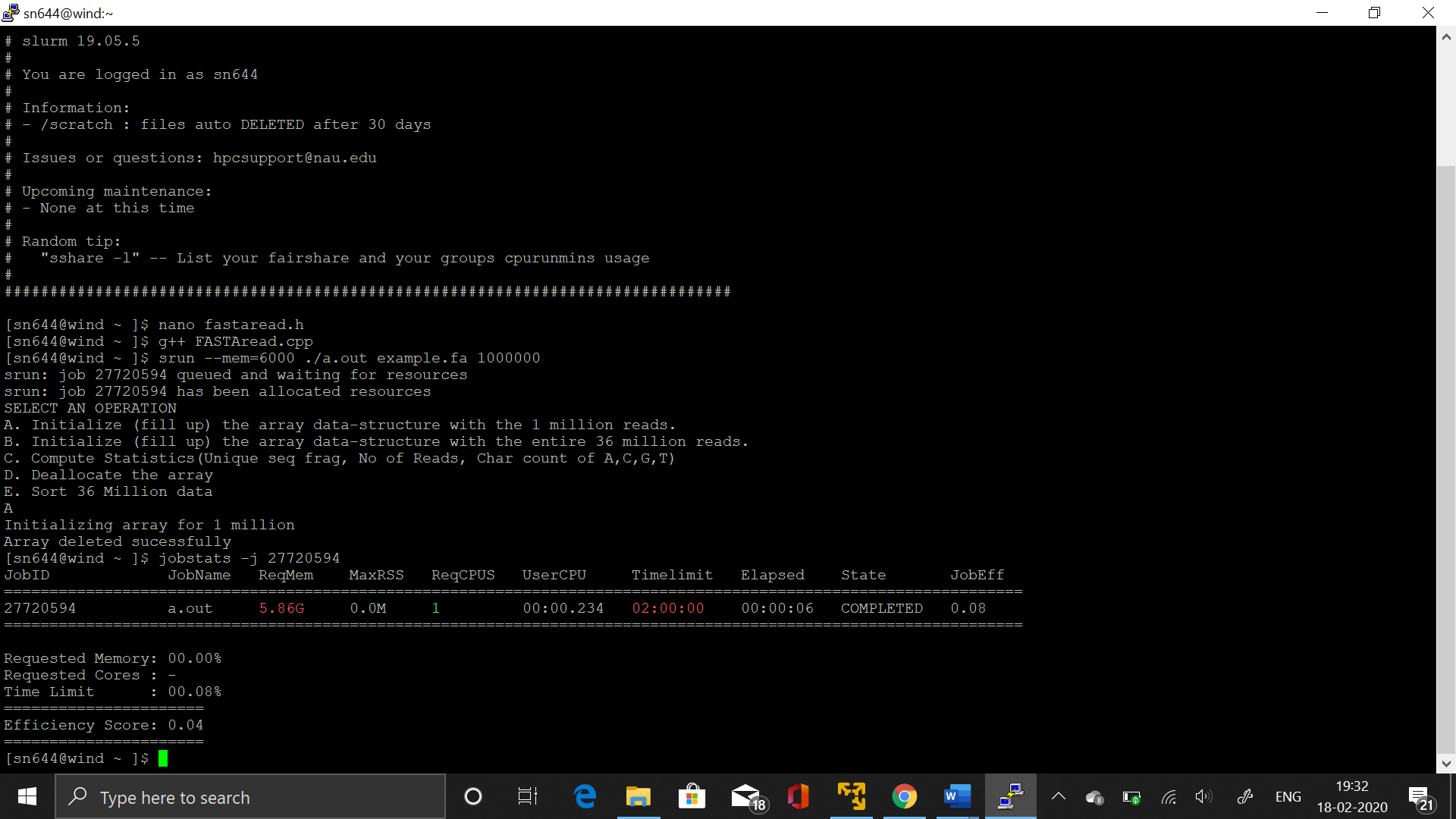
For 1000000 data = 6 seconds

36000000 data = ?

It is 21.6 sec, which is the estimated value.

RAM= **O**(n).

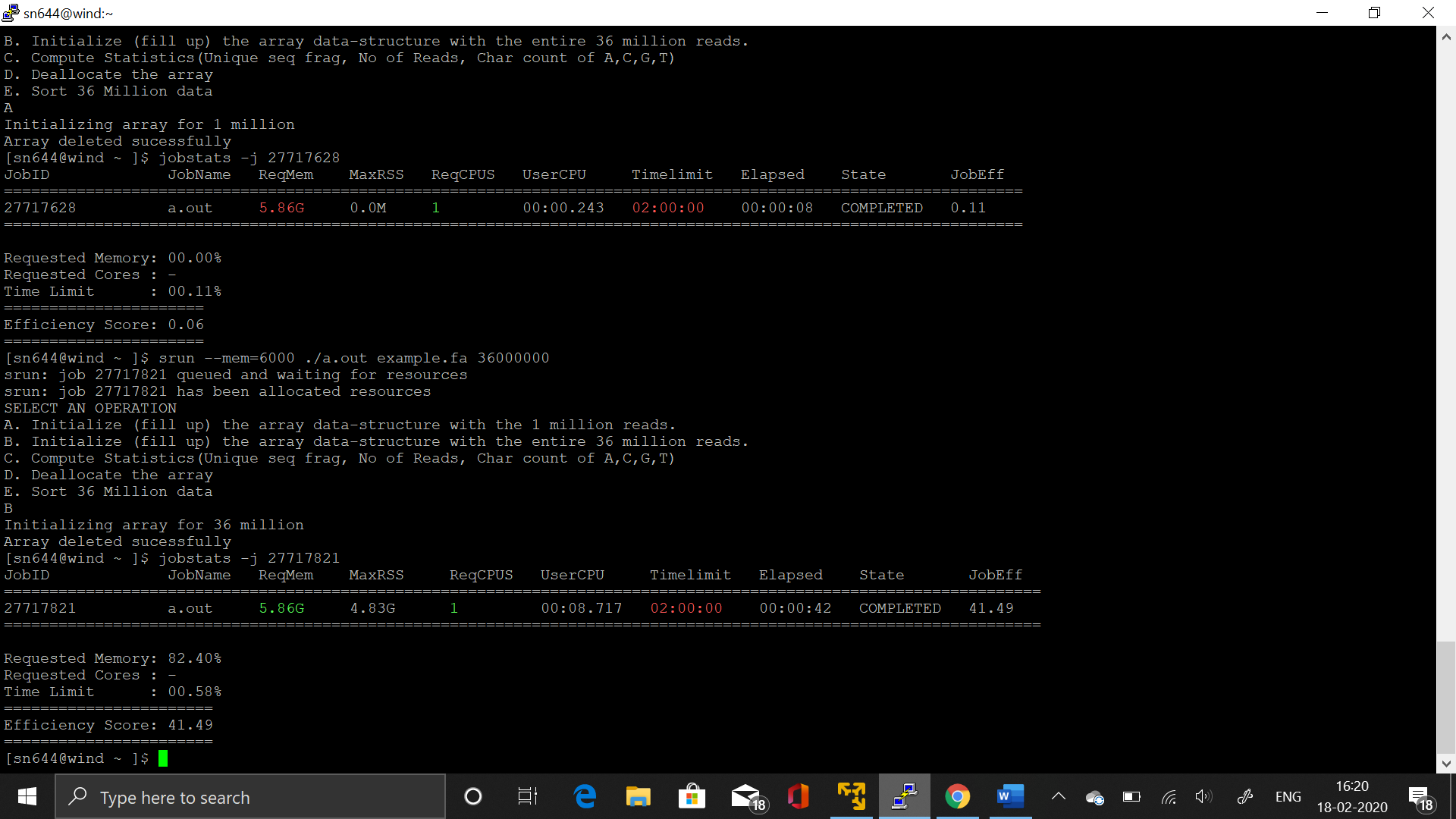
So, For 1 million it takes 3.6GB of RAM



**B.** The following picture tells the actual CPU usage time and RAM it took to initialize 36 Million data .

That is CPU usage = 42 sec

RAM usage=4.83G



The reason for difference in the values for estimated and actual might be because of the efficiency of the job for large dataset.

**C.** Compute Statistics

Output:

**For no of Characters of A,C,G,N.**

A: 493102372

C: 406639890

G: 408544523

N: 489180159

**For number of Reads for 14 columns**

1. 3970133

2. 3790304

3. 3914088

4. 3990231

5. 3970403

6. 3694474

7. 3971901

8. 3969541

9. 3970972

10. 3971121

11. 3976129

12. 3924986

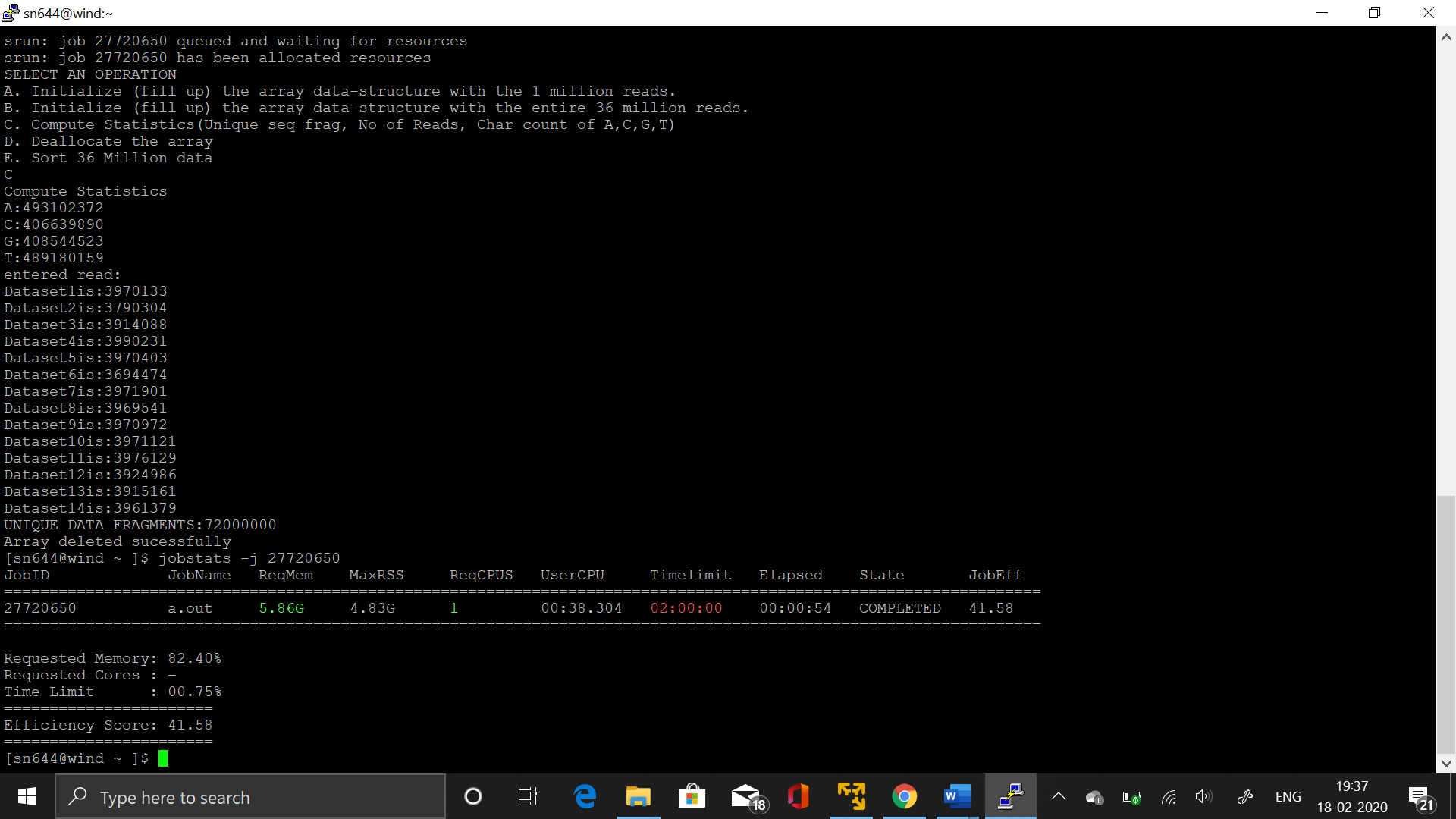
13. 3915161

14. 3961379

**Total number of Unique sequence Fragments**

There are total of 32 million header sets and 32 million Data sequence sets so there must be

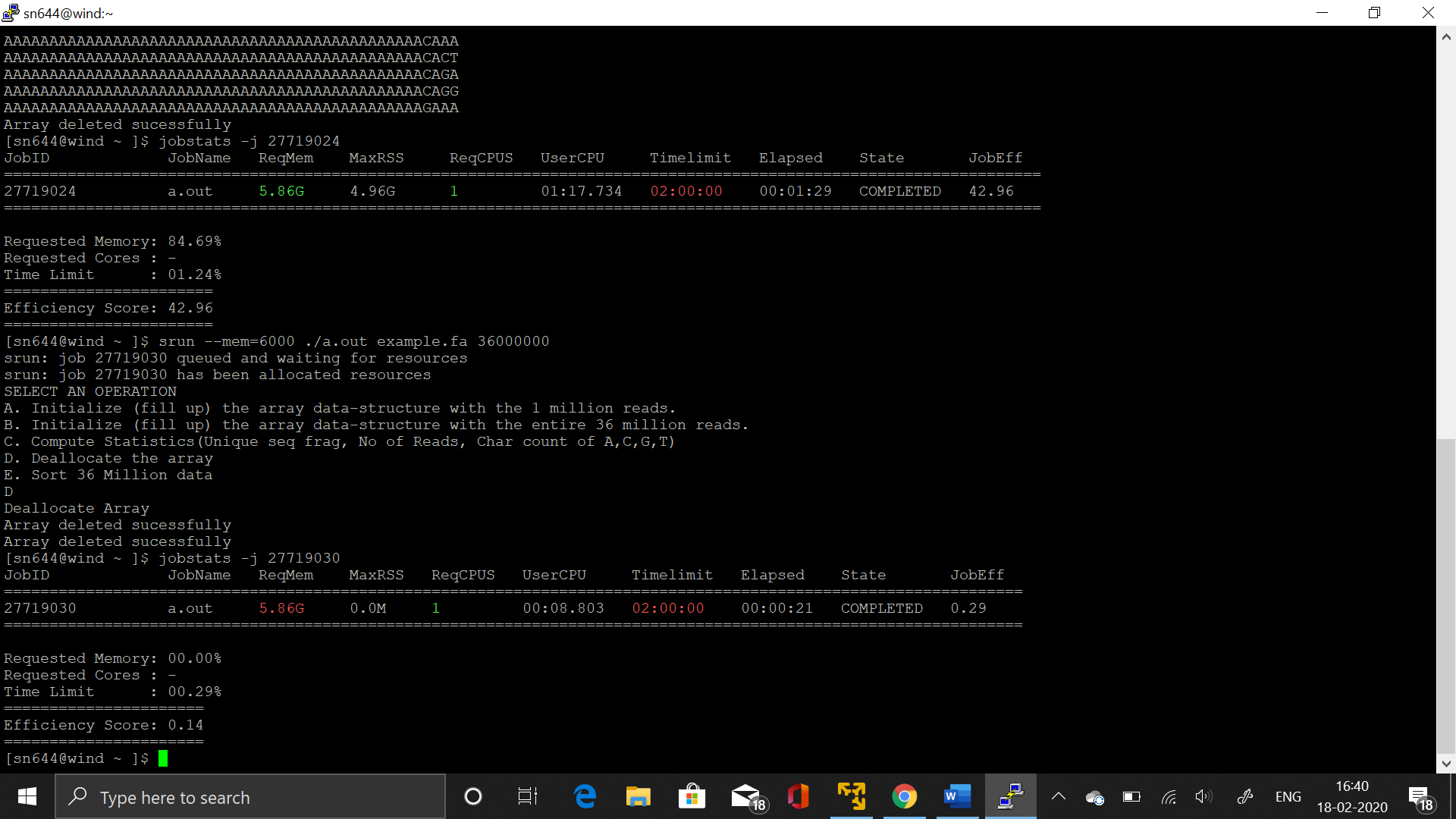
Total of 72 million Unique Sequence Fragments.



**D.** Deallocating the array

~FASTAread()- destructor deallocates the memory as soon as control reaches the end of the class.

On monsoon it took me **“21 seconds”** to deallocate the array.



**E.** Sorting using Merge sort

Time complexity of Merge sort is **O**(nLogn)-which is same in worst case. It’s a logarithmic approach.

Space complexity of Merge sort is **O**(n).

It took RAM = 4.96GB to sort

Time= 1 minute 29 seconds.

