Lab Assignment 3

B21140 / Swarnarup

---

1. **Describe the statement of the problem.**

In this problem we are trying to analyze the performance of the file read throughput of the computer by running simulations. We are trying to look at two factors which are (1)Method of reading by specific C function (2) File size and (3) buffer block size in fread method.

---

2. **Experiment design: List the factors and levels. List the set of experiments you ran, wherein each experiment is defined by the levels chosen for every factor. Justify your choice of experiments.**

Factors and Levels:
   a. Method of Reading: fgetc(), fread()
   b. File Size: 1kb, 10kb, 100kb, 1mb, 10mb, 100mb
   c. Buffer Block Size: 1, 10, 100, 1000, 10000, 100000 (* size of int)

● Full factorial design of File Size and Method of Reading. With 4 reparations. For ANOVA the number of reparations are considered as the number of levels of other factors.

	Methods	
File Size	fgetc	fread
1kb	4 experiments	4 experiments
10kb	4 experiments	4 experiments
100kb	4 experiments	4 experiments
1mb	4 experiments	4 experiments
10mb	4 experiments	4 experiments
100mb	4 experiments	4 experiments

After 4 reparations for each factor across all levels, we are averaging the number. Now when analyzing the factor "methods of reading" we have 2 levels and 6 reparations and when analyzing the factor "file size" we have 6 levels and 2 reparations.

- For analysis of Buffer Block size all levels of File Size and all levels of block size were taken into consideration for the reading method fread.

Table for File Size vs Buffer Block size full factorial

	Throughput in kbps for n Buffer Block Size (n * size of int)					
File Size	1	10	100	1000	10000	100000
1kb	1exp	1exp	1exp	1exp	1exp	1exp
10kb	1exp	1exp	1exp	1exp	1exp	1exp
100kb	1exp	1exp	1exp	1exp	1exp	1exp
1mb	1exp	1exp	1exp	1exp	1exp	1exp
10mb	1exp	1exp	1exp	1exp	1exp	1exp
100mb	1exp	1exp	1exp	1exp	1exp	1exp

Justification:
Since we can't measure the effect of Buffer Block size with respect to the Level 'fgetc' of factor "reading Method" we have taken full factorial of all File size across all Buffer block size with method fread for analysis of the factor 'Buffer Block Size'

Explanation of the Script:

There are 2 C programs fgetc.c and fread.c they compile into fgetc.out and fread.out.
Using a python script we are invoking these .out files to run and the specific file name which has specific size has been passed as argument. The time to run these files are recorded and the throughput is calculated and stored into a csv file.

For analysis of the Buffer block size fread.c is modified and now it's taking another parameter buffer size. Now fread.c compiles to read_with_buffSize.out. Using a python script this .out is passed with various file sizes across various buffer sizes and the throughput data is stored in a csv file.

After the throughput data generation, the analysis and plotting is done in analysis.ipynb

3. **Present the data in the form of tables. Plot the read throughput and discuss your inferences.**

- <u>Analysing File Size and Method of Reading</u>

File Size	Method									
	fgetc				Mean	fread				Mean
1kb	49.3754	66.04147	70.09182	63.54855	62.26431	59.91253	73.99734	79.75118	82.96001	74.155265
10kb	645.3695	805.9317	874.9672	833.3333	789.900425	793.2101	903.0161	819.5378	840.0538	838.95445
100kb	7435.497	7977.663	7242.178	8071.025	7681.59075	6487.609	7121.493	6678.688	8320.16	7151.9875
1mb	46598.41	46649.36	57066.43	48018.76	49583.24	28193.06	28581.8	30094.63	26736.29	28401.445
10mb	113028	110791.3	125608.7	121291.1	117679.775	39683.46	39470.54	34097.42	41399.5	38662.73
100mb	127891.8	127393.3	124914.6	136345.2	129136.225	42367.9	38451.04	38841.64	42745.35	40601.4825

All throughput data is in kb per seconds.

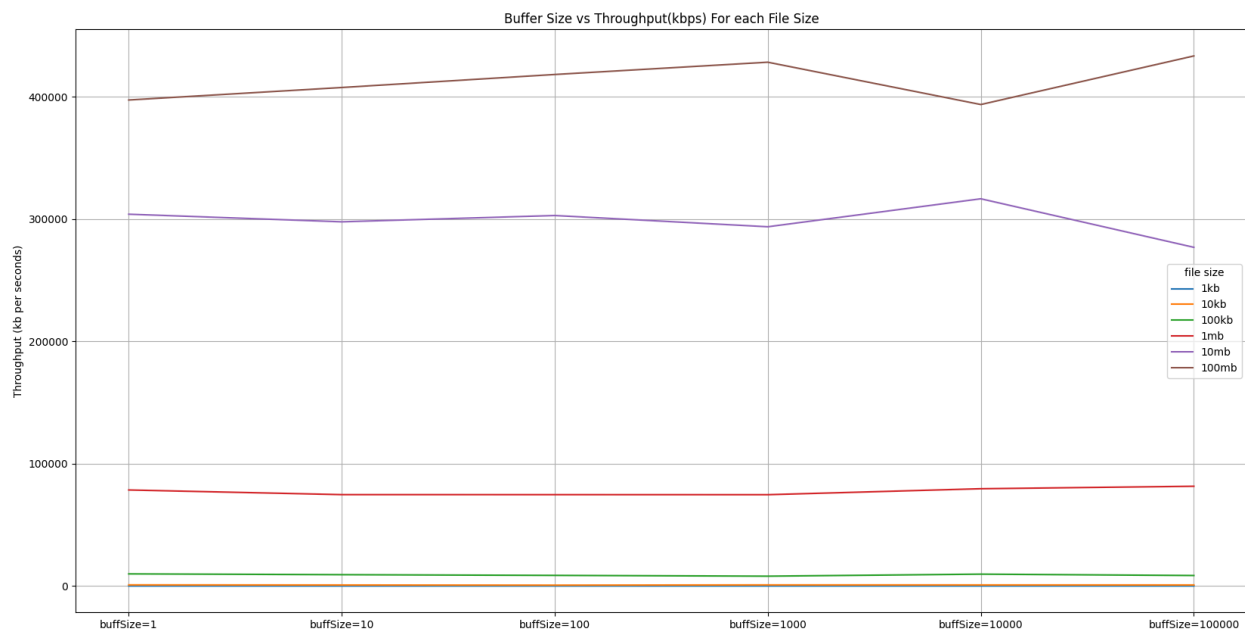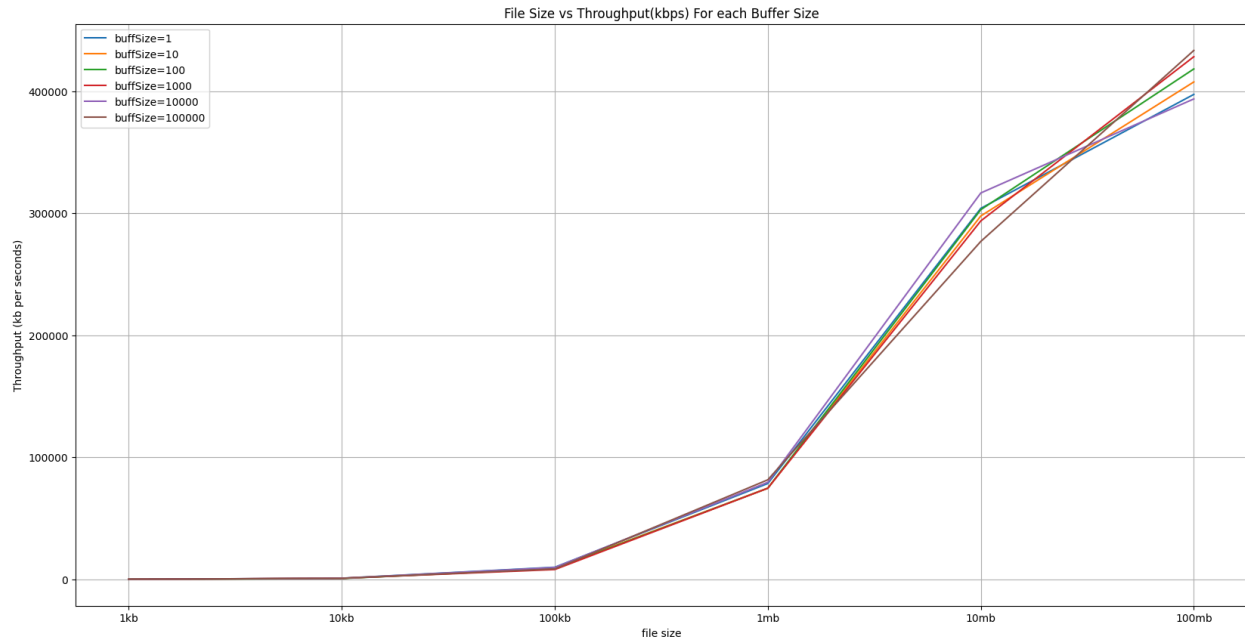File Size vs Throughput(kbps) For each methods
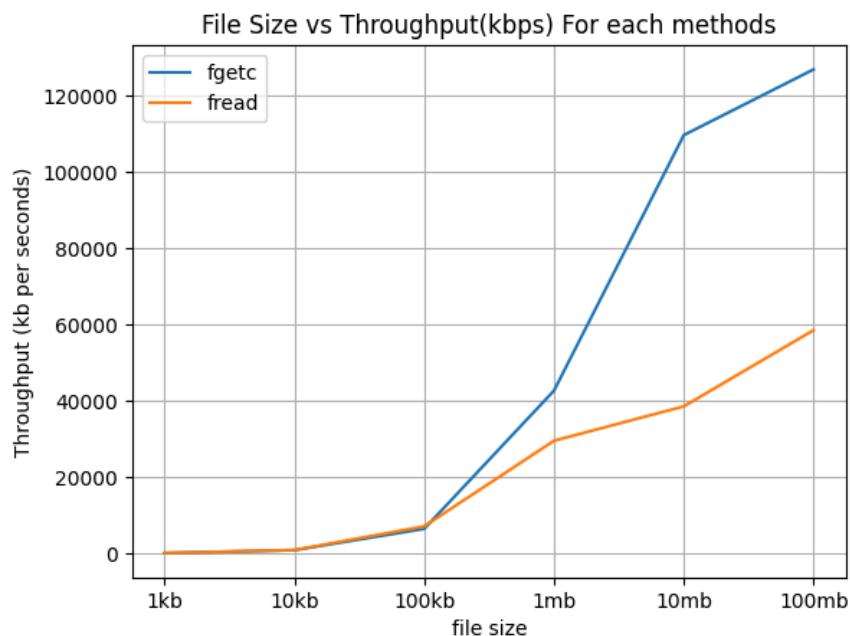
This is a plot for throughput of reading in kbps for different file sizes averaged over all reparations. The two plots show reading throughput in experiment of different file sizes by using different reading methods.

From this plot we can see the throughput increases as the file size increases. Also throughput is sufficiently higher using the "fgetc" method than "fread" method. But it's not obvious whether the factor "reading method" is more influential or the factor "file size".

- ## Analysing File Size and Buffer Block Size

File Size	Throughput in kbps for Different Buffer Block Size (n * size of int)					
	n=1	n=10	n=100	n=1000	n=10000	n=100000
1kb	74.75518	84.01949	80.952	86.38563	85.85165	93.53662
10kb	951.3843	861.6233	742.7765	929.2817	908.7605	880.902
100kb	9953.22	9300.595	8717.636	8081.461	9732.36	8622.92
1mb	78563.76	74755.44	74739.07	74706.35	79583.43	81600.13
10mb	304074.1	297838.9	302994.4	293788.6	316674.9	276981.3
100mb	397453.8	407662.8	418316	428349.7	393737.1	433477.7

Buffer Size vs Throughput(kbps) For each File Size

File Size vs Throughput(kbps) For each Buffer Size

From these 2 Graphs we can conclude that file Size has a Significant effect on the throughput but buffer block size does not have any significant effect on throughput.

4. **Using range method to identify the factor that has the highest influence on the read throughput. Present your analysis in the form of tables. Discuss your findings.**

Analysis of influential factor Using Range Method:

Factor : Method of Reading
Throughput across levels -

Method of Reading	Average Throughput (kb per sec)
fgetc	50822.17
fread	19288.46

Range = Max - Min = 31533.71

Factor : File Size
Throughput across levels -

File Size	Average Throughput (kb per sec)
1kb	68.2097875
10kb	814.4274375
100kb	7416.789125
1mb	38992.3425
10mb	78171.2525
100mb	84868.85375

Range = Max - Min = 84800.641

Factor : Buffer Block Size
Throughput across levels -

Buffer Size	Average Throughput (kb per sec)
1	131845.1699
10	131750.563

100	134265.1391
1000	134323.6297
10000	133453.7337
100000	133609.4148

Range = Max - Min = 2573.066757

>> Hence **Most influential** factor is **File Size**

5. **Using ANOVA analyse the significance of the factor identified above to the variation in the read throughput.**

> **Factor** - Method of Reading

Component	Sum of Squares	% of Variation	Degrees of freedom	Mean Square	F	F-table
y	SSY = 37044615815		ar			
μ = 35055.31252	SSO = a*r*μ^2 = 2 * 6 * 35055.31252 * 35055.31252 = 14,74,64,99,230.498		1			
y - μ	SST = SSY - SSO = 22,29,81,16,584.502		ar-1			
Effect	SSA = 2983123992.699	(SSA/SST)% = 13.37 %	a-1	MSA = SSA/(a-1) = 2983123992.699		
Error	SSE = SST - SSA = 19,31,49,92,591.803	(SSE/SST)% = 86.62 %	a(r-1)	MSE = SSE/a(r-1) = 1,93,14,99,259.18	MSA/MSE = 1.54	0.24

Since MSA/MSE > F-table factor "Method of Reading" is significant.

> **Factor** - File Size

Component	Sum of Squares	% of Variation	Degrees of freedom	Mean Square	F	F-table

y	SSY = 37044615815		ar			
μ	SSO = a*r*μ^2 = 2 * 6 * 35055.31252 * 35055.31252 = 14,74,64,99,230.498		1			
y - μ	SST = SSY - SSO = 22,29,81,16,584.502		ar-1			
Effect	SSA = 15032593839.198	(SSA/SST)% = 67.416 %	a-1	MSA = SSA/(a-1) = 3,00,65,18,767.8396		
Error	SSE = SST - SSA = 7,26,55,22,745.304	(SSE/SST)% = 32.58 %	a(r-1)	MSE = SSE/a(r-1) = 1,21,09,20,457.5506	MSA/MSE = 2.482	0.149

Since MSA/MSE > F-table factor "File Size" is significant.

> Factor - Buffer Block Size

Component	Sum of Squares	% of Variation	Degrees of freedom	Mean Square	F	F-table
y	SSY = 1598270459610		ar			
μ = 133207.9417	SSO = a*r*μ^2 = 6 * 6 * 133207.9417 * 133207.9417 =		1			

	6,38,79,68,06,3 50.2215					
y - μ	SST = SSY - SSO = 9,59,47,36,53, 259.7785		ar-1			
Effect	SSA = 39390720.707	(SSA/SST)% = 0.0041 %	a-1	MSA = SSA/(a-1) = 78,78,144.1 4		
Error	SSE = SST - SSA = 9,59,43,42,62, 539.0715	(SSE/SST)% = 99.995 %	a(r-1)	MSE = SSE/a(r-1) = 31,98,11,42, 084.63	MSA/M SE = 0.00002 4	1.0

Since MSA/MSE < F-table. Hence the factor Buffer Size is insignificant.

>> Since (MSA/MSE)For File Size is Greater than that of Method of Reading, Hence File size is a more Significant factor.