

Análise do impacto de buffers no desempenho da leitura e escrita em arquivos

1 Introdução

O experimento proposto consiste na realização de testes com intuito de determinar a melhora na performance oferecida pela utilização de *buffers* nas operações de entrada e saída em arquivos. Buffers consistem em blocos de memória utilizados para guardar dados temporariamente durante a transferência de uma fonte para um destino, com o intuito de aprimorar a performance de operações onde dados podem ser lidos ou escritos em grupos ao invés de individualmente, como é o caso em discos rígidos.

Para a realização do experimento, foi escrita uma aplicação utilizando a linguagem Java que realiza a criação de arquivos de teste com tamanhos predefinidos e o procedimento de cópia dos mesmos, utilizando diversas configurações distintas, especificadas adiante.

2 O experimento

Inicialmente, foram criados 10 arquivos preenchidos com bytes aleatórios, com tamanho em kilobytes dado pela expressão $\{2^k \mid 6 \leq k \leq 15 \text{ e } k \in \mathbb{N}\}$. Em seguida, foram realizados 50 testes, cada um envolvendo a cópia dos dez arquivos de teste com 4 estratégias distintas para cada arquivo: nenhum buffer, buffer de 1 kB, buffer de 8 kB e buffer de 32 kB. Ao total, foram coletadas 2000 amostras do procedimento de cópia.

O experimento foi realizado em um computador com as seguintes configurações: CPU Intel i3-3240 3.40GHz, RAM 4 GB DDR3 com velocidade de 1067 MHz, HD Samsung HD161HJ com 7200 RPM e sistema operacional Windows 10 Pro 64-bit versão 1511. A plataforma Java utilizada foi o JDK versão 1.8.0 update 66.

2.1 Resultados do experimento e análise

As médias das amostras coletadas foram dispostas na tabela 1, onde a primeira coluna apresenta o tamanho do arquivo em kilobytes e as demais apresentam a média do tempo tomado pelo processo de cópia do arquivo, expresso em milissegundos, separadas de acordo com a estratégia utilizada no procedimento.

Tamanho (kB)	Sem buffer	Buffer de 1kB	Buffer de 8kB	Buffer de 32kB
64	230,82	4,28	3,88	3,92
128	453,74	4,94	4,52	4,34
256	898,18	6,48	5,50	5,34
512	1787,22	9,56	7,50	7,18
1024	3591,98	16,30	12,02	11,16
2048	7162,14	28,44	19,98	19,08
4096	14470,22	54,20	36,46	33,58
8192	28942,04	104,10	72,92	65,92
16384	58018,32	204,30	135,04	125,12
32768	116644,30	407,14	274,12	251,74

Tabela 1. Médias dos tempos de cópia em milissegundos.

As amostras coletadas demonstram claramente que a utilização de buffers no procedimento de cópia oferece um benefício extremamente grande. No mínimo, a melhora foi de 55 vezes, notada na diferença entre a duração de 230,8 ms na cópia sem buffer e 4,2 ms na cópia com buffer de 1 kB, onde o arquivo de teste continha 64 kB. A melhor amostra foi observada na cópia do maior arquivo (32 MB) com o maior buffer (8 kB), onde o processo foi 463 vezes mais veloz.

Os dados apresentados na tabela acima também podem ser verificados visualmente no gráfico 1. Como o tamanho dos arquivos cresce exponencialmente, o eixo Y está em escala logarítmica.

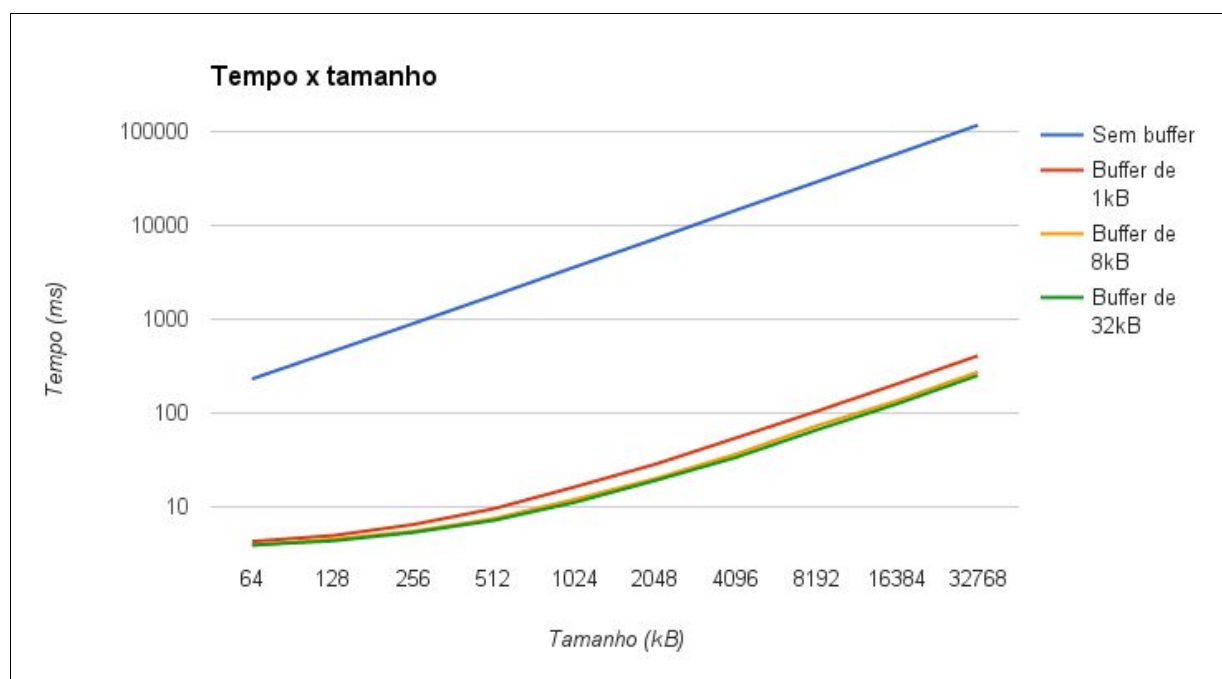


Gráfico 1. Relação entre a duração da cópia e o tamanho do arquivo.

Percebe-se através do gráfico que o tempo de cópia sem buffer teve um crescimento linear em relação ao tamanho dos arquivos. A melhora no desempenho utilizando buffers foi significativa em todos os tamanhos testados, crescendo lentamente até cerca de 4 MB; a partir deste ponto, a melhora se manteve aproximadamente constante em relação à cópia sem buffer.

É também possível perceber que existe pouca diferença entre os tamanhos de buffer, sendo que a diferença no desempenho entre o buffer de 8 kB e o buffer 32 kB é praticamente negligível, havendo uma diferença um pouco maior em comparação com o buffer de 1 kB. Esta característica pode ser observada em detalhes na tabela 2, onde os resultados obtidos foram expressos em megabytes por segundo.

Tamanho (kB)	Sem buffer	Buffer de 1kB	Buffer de 8kB	Buffer de 32kB
64	0,27	14,60	16,11	15,94
128	0,28	25,30	27,65	28,80
256	0,28	38,58	45,45	46,82
512	0,28	52,30	66,67	69,64
1024	0,28	61,35	83,19	89,61
2048	0,28	70,32	100,10	104,82
4096	0,28	73,80	109,71	119,12
8192	0,28	76,85	109,71	121,36
16384	0,28	78,32	118,48	127,88
32768	0,27	78,60	116,74	127,12

Tabela 2. Taxa de transferência expressa em megabytes por segundo.

Observando a tabela, é imediatamente perceptível que o desempenho da cópia sem buffer é bastante ineficaz, não atingindo nem mesmo 1 MB/s, e apresentando em média uma taxa de transferência de 270 kB/s. A mudança pequena a partir do tamanho de 4 MB é evidenciada pelas quatro últimas linhas da tabela, onde a melhora se mantém pequena em todas as amostras.

Os dados apresentados na tabela acima foram representados de forma visual no gráfico 2, onde o crescimento gradual da velocidade é claramente visível. É também possível notar a diferença ínfima entre o buffer de 8 kB e o buffer de 32 kB.

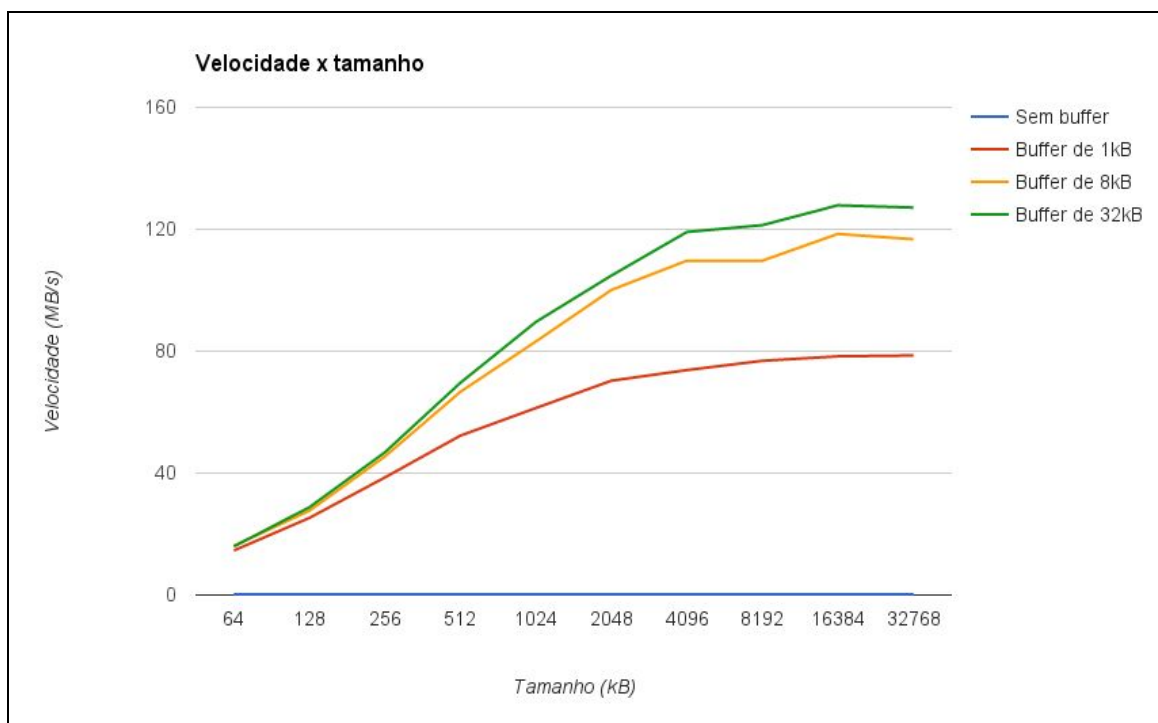


Gráfico 2. Performance da cópia em função do tamanho do arquivo.

O desvio padrão relativo é apresentado no gráfico 3. O desvio padrão de cada conjunto de dados foi dividido por sua respectiva média, obtendo assim o desvio relativo, expresso em porcentagem.

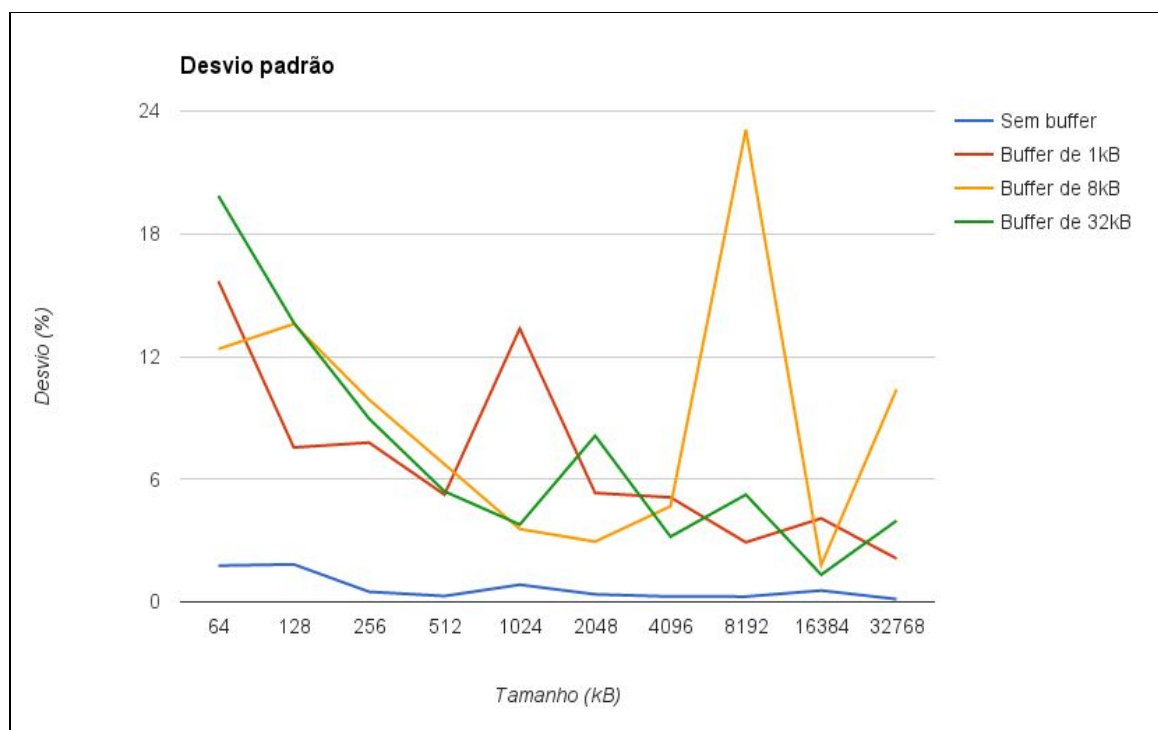


Gráfico 3. Desvio padrão relativo das amostras.

É possível observar um comportamento bastante caótico quando a cópia é realizada com buffers, independentemente do tamanho do buffer; o mesmo comportamento não é apresentado pela cópia sem buffer, onde o desvio padrão relativo se manteve pequeno e aproximadamente constante em todas as amostras coletadas.

Sendo assim, uma possível hipótese levantada para tal resultado é a fragmentação do disco rígido, pois como a escrita dos dados está sendo realizada em blocos ao invés de byte por byte, o tempo necessário para acomodar tais blocos pode variar bastante se houverem poucos setores livres contíguos. Para comprovar ou rejeitar a hipótese, o experimento deveria ser realizado mais uma vez após desfragmentar o disco rígido e os dados deveriam ser analisados novamente, verificando se o comportamento se repete ou não.

3 Conclusões

Observando os dados coletados no experimento, foi possível concluir que a operação sem buffer apresentou performance substancialmente ineficiente, sendo mais lenta do que boa parte das taxas de download de Internet domiciliar no Brasil. Todas as operações com buffer foram significativamente superiores, atingindo em média um desempenho 266 vezes mais veloz. Os tamanhos de buffer analisados apresentaram velocidade similar, com o maior tamanho testado exibindo apenas cerca de 15% de melhora em relação à média.

Todos estes benefícios não apresentaram nenhuma desvantagem perceptível, especialmente na implementação do programa, escrito na linguagem Java. Grande parte do algoritmo utilizado no procedimento sem utilização de buffers pôde ser reaproveitado, sendo necessário apenas alterar o subtipo do mecanismo de transferência, pois ambos os casos utilizam a mesma superclasse, permitindo a aplicação do Princípio de Substituição de Liskov.