

Um Estudo Experimental de Avaliação de Armazenamento em Memória Secundária com Foco em Alto Desempenho

Fernando de Almeida Silva¹, Mario Antonio Ribeiro Dantas¹

¹Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG – Brazil

fernandoalmeida.silva@ufjf.br, mario.dantas@ice.ufjf.br

Abstract. *This research work presents an experimental effort that aims to elaborate and present a study focusing on the modicity of the cost for data transfer and storage in secondary storage device, focusing on the use of fog layers, with a focus on high performance. Our initial experimental results illustrate the technical feasibility of the proposal and its differential effect, including on computers that would not be considered adequate computational environments.*

Resumo. *Este trabalho de pesquisa apresenta um esforço experimental que tem por objetivo elaborar e apresentar um estudo com foco na modicidade do custo para transferência e armazenamento de dados em dispositivo de armazenamento secundário, focando na utilização nas camadas de fog, com foco em alto desempenho. Nossos resultados experimentais iniciais ilustram a viabilidade técnica da proposta e o efeito diferencial da mesma, inclusive em computadores que não seriam considerados como ambientes computacionais adequados.*

1. Introdução

O desenvolvimento tecnológico permitiu o uso cada vez mais amplo da tecnologia Internet das Coisas (*Internet of Things – IoT*) nos mais diversos ambientes e cenários, desde de ambientes industriais até monitoramento ambiental e o controle da saúde humana, por exemplo. Tal cenário tem ocasionado um aumento constante na quantidade de dados gerados, consequentemente um crescimento na necessidade de processamento e armazenamento e que aumentará nos próximos anos.

Segundo Balakrishna e Vijender (2020), no cenário atual, cerca de 35 bilhões de dispositivos *IoT* estão conectados à Internet. Sendo que até 2025, a previsão é que o número cresça entre 80 e 120 bilhões de dispositivos, gerando por ano cerca de 180 trilhões de gigabytes de novos dados por meio desses dispositivos.

Tal cenário evidencia a camada *Fog* entre as camadas *Edge* e a *Cloud* para trazer o armazenamento mais perto dos nós *Edge*, organizando e pré-processando os dados oriundos desses dispositivos, permitindo maior escalabilidade com o mínimo de impacto na rede de dados. Entretanto nesse contexto a quantidade e qualidade limitada da memória secundária pode ser um impedimento.

Diante disso, a contribuição deste trabalho está em apresentar um estudo preliminar de desempenho sobre técnicas e dispositivos de armazenamento em memória

secundária comparando com o desempenho da memória principal convencional, analisando assim a taxa de transferência de dados usando tecnologias atuais empregadas em clusters de HPC ou data centers. Além disso, que tenham um custo acessível para uso em dispositivos da camada *Fog* e estendendo a vida útil de equipamentos já disponíveis e conseqüentemente minimizando o impacto ambiental dessa proposta.

Este trabalho está estruturado conforme descrito a seguir. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 apresenta a metodologia desenvolvida nesse estudo. O ambiente utilizado e os resultados obtidos são apresentados na Seção 4. Por fim, a Seção 5 faz a conclusão e se propõe os trabalhos futuros.

2. Trabalhos Relacionados

Uma arquitetura de computação *Fog* tem como objetivo melhorar o gerenciamento de rede, o armazenamento e o processamento de aplicativos. Para esse propósito, as arquiteturas *Fog* integram mecanismos que podem distribuir melhor os recursos em uma infraestrutura específica (Silva *et al*, 2019).

Entre as principais fontes de dados digitais (Coughlin, 2017) temos medicina, geociências e outras áreas. Isso resulta no crescimento do armazenamento, tanto local, bem como do armazenamento na nuvem e na camada *Fog* entre os nós *Edge*. Para retenção de longo prazo, precisaremos de armazenamento em massa barato e tecnologias de processamento e memória mais rápidas para analisar e usar dados (Habibi, 2020).

Entre as opções usualmente existentes de armazenamento, podemos citar o *Hard Disk Drive* (HDD) que pode armazenar uma grande quantidade de dados (alta capacidade) e o preço de um HDD é usualmente mais barato em relação ao valor monetário por bytes de armazenamento. Porém tem ganhado mais espaço no mercado como dispositivo de armazenamento o *Solid State Drives* (SSD). Pois entre as vantagens, o mesmo é um dispositivo de armazenamento não mecânico e não magnético em que os dados são armazenados usando circuitos eletrônicos, assim são menos suscetíveis a choques físicos, mais silenciosos e têm desempenho geral melhor do que o de HDD. Por exemplo, levaria mais de 15 segundos para ler uma imagem de ressonância magnética de 1,8 GB armazenada em um HDD em comparação apenas 0,15 segundos num dispositivo de armazenamento SSD. Entretanto por utilizar a interface SATA III, a largura de banda efetiva atinge o pico em cerca de 555 MB/s, limitando assim velocidade de acesso e armazenamento do SSD (Armoogum e Khonje, 2021).

3. Metodologia do Estudo Experimental

Neste trabalho optou-se por se fazer um estudo e levantamento das tecnologias e dispositivos de armazenamento em memória secundária das últimas décadas. Além disso, avaliar as vantagens e desvantagens de cada um e como poderiam contribuir numa proposta mais eficiente.

Realizamos um levantamento de pesquisas relacionadas à avaliação de desempenho de E/S em Computação de Alta Performance entre 2015 e 2021. Identificamos artigos base levando em conta o resumo e a conclusão considerando principalmente os que abordam o mesmo tipo de pesquisa, desafios e referências comuns a proposta, conforme trabalho de pesquisa indicado em Pioli (2020).

Alcançando essa proposta teórica, o próximo passo foi implementar a mesma criando um exemplo prático, que poderia ser implementado inclusive em computadores que não teriam nativamente a conexão necessária, permitindo assim estender a vida útil desses equipamentos minimizando o impacto ambiental. Pois segundo a Organização das Nações Unidas (ONU) globalmente por ano são gerados aproximadamente 50 milhões de toneladas de lixo eletrônico conhecido como *e-waste* (Thakur e Kumar, 2021).

Em prosseguimento testar, analisar e considerar os resultados obtidos, verificando a viabilidade da proposta e o impacto dela nos sistemas atuais. Assim como, a mesma pode ser melhorada em pesquisas futuras.

4. Ambiente e Resultados Experimentais

Neste projeto foi utilizado um conjunto de hardware e software para implementar e avaliar a viabilidade da proposta em diferentes gerações do barramento PCI Express, por exemplo, a segunda geração do barramento que foi utilizada no experimento. Assim, verificando a possibilidade de retrocompatibilidade, permitindo o uso de computadores atuais e outros modelos considerados obsoletos, por exemplo, que não possuem um conector nativo para dispositivos Non-Volatile Memory express (NVMe).

4.1 Ambiente

Para simular a viabilidade da proposta, foi utilizado no experimento um computador com uma década de uso tendo um processador Intel® Core i5-2500 CPU @ 3.30GHz, placa-mãe Intel modelo DQ67SW e 8GB de RAM (memória principal). Complementam o hardware de forma relevante, o dispositivo NVMe P400 da marca Reletech e uma placa de expansão NVMe para PCI Express para slot x16.

Entre os softwares para implementar a proposta, temos o Sistema Operacional Microsoft Windows 10 Pro (64 bits). Além desse, semelhante (Clark, 2017) foi utilizado da empresa Dataram o software RAMDisk, que permite utilizar uma parte da memória principal do sistema como unidade de disco virtual. Sendo esse uma referência de comparação para os diferentes dispositivos de armazenamento secundários. A versão utilizada foi a 4.4.0.36 (freeware) que permite criar discos virtuais de até 1 GB de tamanho.

Na bateria de testes de leitura e escrita em diferentes cenários, semelhante (Clark, 2017) o software utilizado foi o *Crystal Disk Mark*, sendo a versão utilizada a 3.0.4 x64, pois é a versão mais recente compatível com partições de 1 GB de tamanho ou menor, que é a restrição do software RAMDisk na versão freeware. O mesmo realiza quatro tipos de testes: Seq, 512 K, 4K e 4K QD32, sendo: Seq com blocos sequenciais de 1024 KB; 512 K com blocos de 512 KB, gravados em locais aleatórios; 4 K os testes mantém uma fila de gravação e leitura com um bloco de 4 KB; 4 K QD32 os testes mantém uma fila de gravação e leitura com 32 blocos de 4 KB.

4.2 Resultados Experimentais

No experimento utilizando o software *Crystal Disk Mark*, conforme figuras 1 e 2, temos um comparativo de leitura e escrita em diferentes cenários entre dois dispositivos de

memória secundária, a esquerda um HD com partes móveis e a direita um SSD ambos tendo conexão SATA.

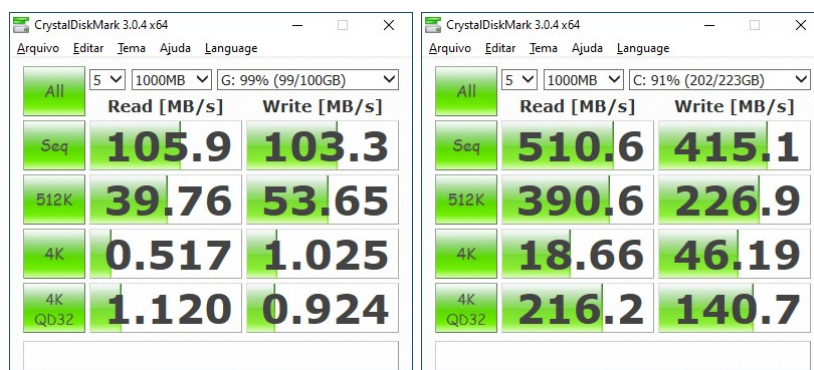


Figura 1. Comparativo de leitura e escrita entre HD e SSD.

Continuando o experimento temos um comparativo de leitura e escrita em diferentes cenários entre o dispositivo de memória secundária NVMe a esquerda e um HD virtual criado pelo software RAMDisk que seria o limite máximo teórico.

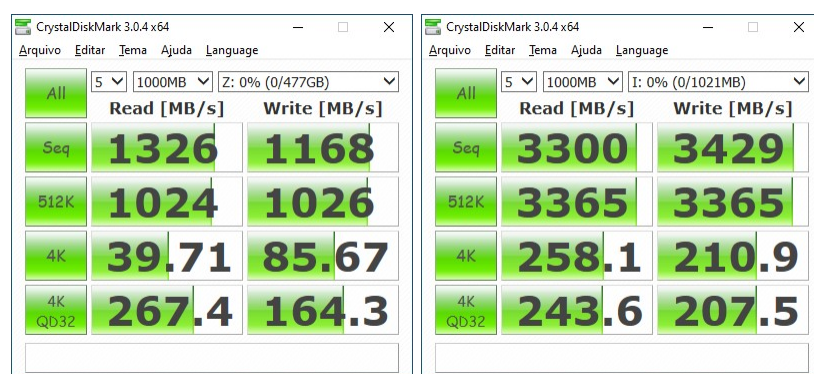


Figura 2. Comparativo de leitura e escrita entre NVMe e HD virtual.

Como resultado dos experimentos, foram elaboradas comparações ilustradas na tabela 1. Calculando o percentual de eficiência dos dispositivos de memória secundária, ou seja, as taxas de leitura e gravação de cada dispositivo de armazenamento secundário em relação HD virtual.

Tabela 1. Percentual comparativo entre memórias secundárias e a virtual.

Tipo de teste:	HD	SSD	NVMe	Leitura HD virtual (MB/s)	HD	SSD	NVMe	Escrita HD virtual (MB/s)
Seq	3%	15%	40%	3300	3%	13%	35%	3429
512K	1%	12%	31%	3365	2%	7%	31%	3365
4K	0,02%	1%	1%	258	0,03%	1%	3%	211
4K QD32	0,03%	7%	8%	244	0,03%	4%	5%	207

5. Conclusões e Trabalhos Futuros

Os experimentos apresentados neste trabalho indicam a viabilidade da proposta cujo objetivo foi prover uma melhoria na forma de armazenamento considerando os

dispositivos HDD, SSD e NVMe, inclusive em computadores que não têm uma interface para dispositivos NVMe. Essa abordagem foi atingida por meio de adaptadores que utilizam slots PCI Express (PCIe) disponíveis na placa mãe. Tendo em vista os resultados apresentados, o uso dessa proposta viabiliza um armazenamento com características de suporte para alto desempenho, inclusive em computadores que não seriam considerados como ambientes computacionais adequados. Verificando inclusive a compatibilidade em diferentes gerações da interface PCI Express. Obtendo taxas de leitura e escrita superiores as de HDD e SSD, contornando as limitações da interface SATA III e utilizando a ponte norte, da mesma maneira que a memória principal.

Como trabalhos futuros, outros ambientes também deverão ser avaliados, considerando a diversificação da aplicação. Por exemplo, estudar formas de paralelismo, permitindo a leitura e gravação simultânea de vários dispositivos NVMe, além de mensurar o desempenho e o consumo energético em diferentes cenários. Avaliar computadores com diferentes gerações de barramento PCI Express e outras configurações. Em adição, efetuar uma pesquisa visando o desenvolvimento de soluções de software que utilizam essa memória secundária em seus melhores cenários de armazenamento, como por exemplo, em conjunto com a memória principal. Pretende-se investigar em futura avaliação a relação de perda de desempenho associada ao armazenamento na memória principal convencional.

Referências

- Balakrishna, Sivadi, M. Thirumaran, and Vijender Kumar Solanki. "IoT sensor data integration in healthcare using semantics and machine learning approaches." *A Handbook of Internet of Things in Biomedical and Cyber Physical System*. Springer, Cham, 2020. 275-300.
- Silva, Daniel Maniglia A. da, et al. "An analysis of fog computing data placement algorithms." *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 2019.
- Coughlin, Tom. "A Solid-State Future [The Art of Storage]." *IEEE Consumer Electronics Magazine* 7.1, 2017. 113-116.
- Habibi, Pooyan, et al. "Fog computing: a comprehensive architectural survey." *IEEE Access* 8 (2020): 69105-69133.
- Armoogum, Sandhya, and Patricia Khonje. "Healthcare Data Storage Options Using Cloud." *The Fusion of Internet of Things, Artificial Intelligence, and Cloud Computing in Health Care*. Springer, Cham, 2021. 25-46.
- Pioli, Laercio, et al. "Characterization Research on I/O Improvements Targeting DISC and HPC Applications." *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2020.
- Thakur, P., and S. Kumar. "Evaluation of e-waste status, management strategies, and legislations." *International Journal of Environmental Science and Technology*, 2021. 1-10.
- Clark, Christopher J. *Planning for the Influence of Emerging Disruptive Technologies on IT Systems*. Diss. The George Washington University, 2017.