

Relatório de Desempenho do Novo SSD de Alta Capacidade para Aplicações 3D

coleta e interpretação de dados

New High Capacity SSD Performance Report for 3D Applications

data collection and interpretation

B. A. F. S. Arruda, D. D. Melo

Universidade Federal de Pernambuco – Centro de Informática

Recife, Brasil

bafsa; ddm / @cin.ufpe.br

Resumo - Neste artigo, é apresentada uma avaliação detalhada de um novo SSD de alta velocidade destinado a melhorar o desempenho de transferência de dados em empresas de desenvolvimento 3D. Este SSD promete duplicar a maior capacidade atual de leitura no mercado, essa leitura de dados é fundamental para lidar com os grandes volumes de dados gerados por animações, modelos e renderizações 3D. A pesquisa incluiu testes de hipótese para verificar se o conjunto de dados se trata de uma distribuição normal e para verificar se o desempenho do SSD atende à meta declarada de 29.000 MB/s, fazendo uso do teste U de Mann-Whitney. No entanto, os resultados revelaram que o SSD não alcançou essa promessa, levantando questões sobre sua prontidão para uso empresarial.

Palavras Chave - SSD de alta velocidade; desenvolvimento 3D; transferência de dados; desempenho.

Abstract - In this article, a detailed evaluation of a new high-speed SSD aimed at improving data transfer performance in 3D development firms is presented. This SSD promises to double the current highest reading capacity in the market, which is crucial for handling large volumes of data generated by 3D animations, models, and renderings. The research included hypothesis testing to verify if the dataset follows a normal distribution and to determine if the SSD performance meets the stated goal of 29,000 MB/s, using the Mann-Whitney U test.

However, the results revealed that the SSD did not achieve this promise, raising questions about its readiness for enterprise use.

Keywords - High-speed SSD; 3D development; data transfer; performance.

I. Introdução

A transferência rápida de dados desempenha um papel vital em todas as áreas do mercado atual. Desde a computação pessoal até as aplicações empresariais e industriais, a capacidade de mover grandes quantidades de dados de forma eficiente não apenas aumenta a produtividade, mas também impulsiona a inovação e abre novas possibilidades em termos de análise de dados em tempo real, aprendizado de máquina e automação.

No contexto específico das empresas de desenvolvimento 3D, onde arquivos de projeto podem chegar a gigabytes de tamanho, a velocidade e a eficiência do armazenamento têm um impacto direto na capacidade de colaboração em equipe, na velocidade de processamento de renders e na capacidade de atender prazos apertados. Um SSD capaz de lidar com essas demandas de forma eficaz não só melhora a qualidade do trabalho produzido, mas também permite que as empresas entreguem projetos de forma mais rápida e econômica.

A indústria de desenvolvimento 3D está constantemente evoluindo, com empresas que produzem animações, modelos e renderizações enfrentando desafios crescentes relacionados ao

tamanho e complexidade dos arquivos. Em resposta a essas demandas, o desenvolvimento de tecnologias de armazenamento de dados mais rápidas e eficientes é crucial para manter a competitividade e a eficiência operacional. Neste contexto, a avaliação de um novo SSD que promete dobrar o recorde de velocidade de leitura para 29.000 MB/s é de grande relevância para toda a indústria.

II. Metodologia

A. Configuração do setup

Para garantir o objetivo da pesquisa de atestar a real capacidade do SSD, foi montado um computador de forma que seus componentes não afetassem a investigação da peça avaliada. Assim, o setup utilizado tem os seguintes componentes:

GPU ASUS ROG Strix GeForce RTX 4090 OC Edition, CPU Intel Core i9-13900K, CPU Cooler Noctua NH-D15 Chromax Black, RAM Corsair Vengeance LPX 32 GB (2x16), Motherboard ASUS ROG Strix Z790-E, Lian Li PC-O11 Dynamic, Fonte Corsair RM1000x 1000W 80+ Gold, Elgato 4K60 Pro Mk.2

B. Coleta dos dados

Para avaliar a eficácia do novo SSD, realizamos um estudo detalhado. Registramos continuamente os dados de velocidade de transferência a cada segundo diretamente na máquina de teste, totalizando 1000 observações ao longo de 16 minutos e 40 segundos. Os dados foram então submetidos a quatro testes de hipótese, divididos em duas sessões distintas.

C. Bibliotecas utilizadas na análise dos dados

Durante a realização do nosso trabalho, nossos dados foram avaliados enquanto faziam uso da linguagem Python. Também foi feito o uso das bibliotecas Matplotlib, Seaborn e Spicy para a geração dos gráficos, enquanto para os cálculos das informações tanto NumPy quanto Statistics foram utilizadas.

D. Teste de hipótese I - Distribuição normal

Para verificar a normalidade dos dados coletados, utilizamos duas abordagens robustas: o gráfico Q-Q (quantile-quantile plot) e o teste *normaltest* da biblioteca Scipy. O gráfico Q-Q proporcionou uma visualização imediata da conformidade dos dados com uma distribuição normal, enquanto o teste *normaltest* aplicou uma análise estatística mais formal. Este último avalia a curtose e a assimetria dos dados para determinar se a distribuição é normal.

Estabelecemos a Hipótese Nula (H_0) de que os dados seguem uma distribuição normal, enquanto com a Hipótese Alternativa (H_a) consideramos que os dados não seguem uma distribuição normal. A escolha do teste de D'Agostino e Pearson na função *normaltest* da Scipy se justifica pela sua capacidade de lidar com diferentes formas de distribuição. Essa abordagem metodológica combinada foi essencial para validar estatisticamente a normalidade dos dados, assegurando a confiabilidade das conclusões alcançadas.

E. Teste de hipótese II - Eficiência do SSD

Para analisar o desempenho médio do SSD, utilizamos o teste estatístico não paramétrico Wilcoxon-Mann-Whitney através da função *wilcoxon* da biblioteca Scipy. Estabelecemos duas hipóteses: na Hipótese Nula (H_0), assumimos que a média da velocidade do SSD é 25.000 MB/s ($\mu = 25.000$), enquanto na Hipótese Alternativa consideramos que a média real do SSD difere significativamente desse valor, o que iria contra o resultado esperado. Após realizar a análise estatística, encontramos evidências robustas para concretizar o resultado dos dados analisados a fim de ditar se o novo produto atende as expectativas.

III. Resultados e Discussão

F. Análise descritiva dos dados

Os dados coletados representam a quantidade contínua de dados transferidos por segundo pelo novo SSD, variando no intervalo [17,528.8, 28,721.8]. Para explorar a tendência central desses dados, foram computadas medidas de centralidade e dispersão, fundamentais para compreender como os valores se agrupam em torno de pontos médios e quanto variam em relação a esses valores centrais. Esses resultados foram detalhadamente apresentados na Tabela 1, oferecendo uma visão abrangente da distribuição dos dados e destacando tanto a média quanto a variabilidade das quantidades de dados transferidos.

TABELA 1. Medidas centrais de dispersão

Medidas	Resultados
Média	23132.16
Mediana	23193.60
Moda	25896.50
Desvio padrão	3243.94
Variância	10523159.14
Coefficiente de variância	0.14
1º Quartil	20293.77
2º Quartil	23193.60
3º Quartil	25870.52
4º Quartil	28721.80
Curtose	-1.17

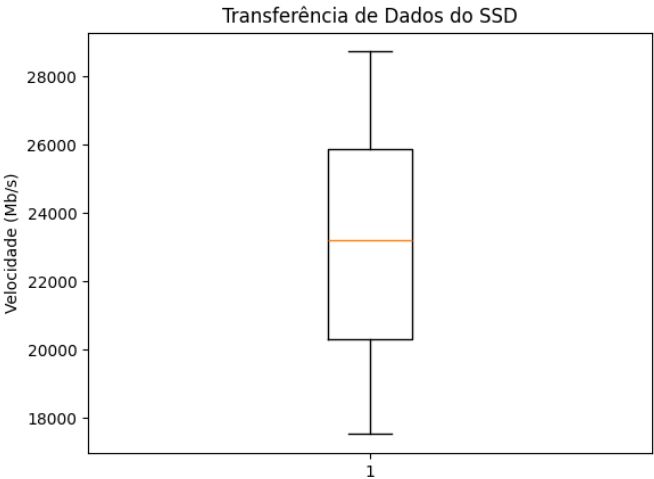
Com base nos dados estatísticos apresentados, podemos interpretar diversos aspectos importantes da distribuição dos dados. A média (23132.16) indica o ponto central ao redor do qual os dados estão agrupados, refletindo o valor médio da amostra analisada. Por outro lado, a mediana (23193.60) é crucial para representações gráficas como o boxplot, dividindo os dados exatamente ao meio e indicando a posição central da distribuição. A moda (25896.50) é significativa em gráficos de barras ou histogramas, mostrando onde ocorre o pico mais alto na distribuição dos dados. O desvio padrão (3243.94) revela a magnitude da dispersão dos dados em relação à média, indicando uma considerável variabilidade dos valores em torno do valor médio.

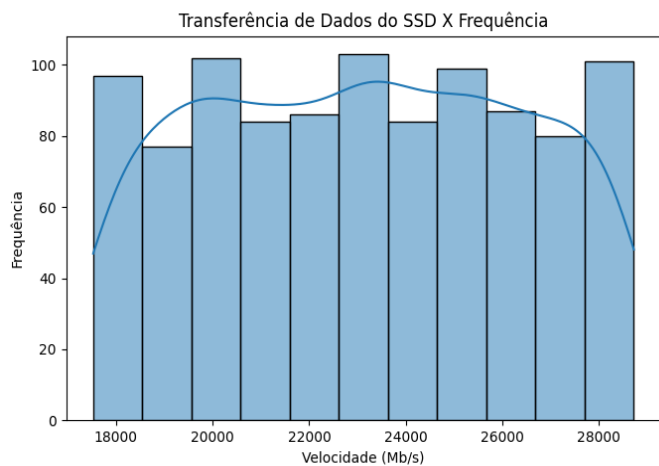
Além disso, a variância (10523159.14), que é calculada ao quadrado do desvio padrão, amplia essa percepção de dispersão, enfatizando valores mais distantes da média. Por outro lado, o coeficiente de variância (0.14) expressa essa variabilidade como uma porcentagem da média, mostrando que, apesar da dispersão, ela é moderada em relação ao valor médio. Os quartis (1º: 20293.77, 2º: 23193.60, 3º: 25870.52, 4º:

28721.80) dividem os dados em quatro partes iguais, úteis para representações gráficas como o boxplot, fornecendo insights sobre a distribuição dos valores ao longo do conjunto de dados. Finalmente, a curtose (-1.17) indica que a distribuição dos dados é mais achatada e possui caudas menos pesadas do que uma distribuição normal padrão, oferecendo uma visão adicional sobre a forma da distribuição observada.

Utilizando as medidas de centralidade e de dispersão, foram elaborados os gráficos de boxplot (Figura 1) e histograma (Figura 2). Além disso, desenvolvemos um gráfico temporal mostrando a quantidade de dados transportados por um SSD em Mb/s ao longo do tempo de observação (Figura 3).

Esses gráficos nos fornecem uma análise detalhada da distribuição dos dados de velocidade de transferência do SSD, permitindo identificar tanto a tendência central como a variabilidade dos valores observados. O boxplot revela a mediana, quartis e possíveis outliers na distribuição dos dados, enquanto o histograma oferece uma visão da frequência com que diferentes faixas de velocidade de transferência ocorrem. Por fim, o gráfico temporal nos ajuda a entender qualquer padrão ou variação ao longo do tempo na performance do SSD em termos de velocidade de transferência.



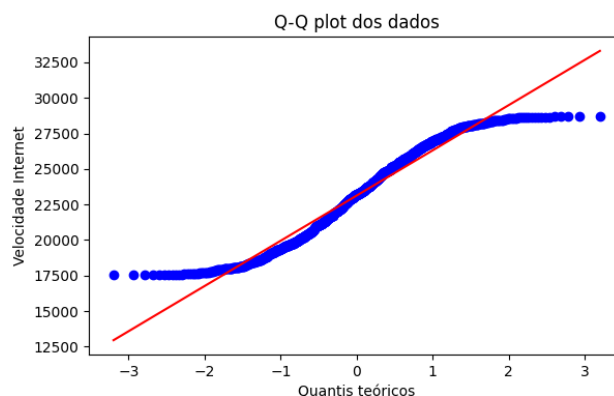


A análise da Figura 1 revela que os dados relacionados à velocidade de transferência do SSD exibem uma distribuição que se aproxima da simetria em torno da mediana. Não foram identificados outliers significativos que necessitassem ser excluídos do conjunto de dados, uma vez que todos os valores observados estão dentro do intervalo esperado para essa métrica. É importante notar que variações na velocidade de transferência são comuns e naturalmente esperadas, influenciadas por diversos fatores, como a complexidade das operações realizadas pelo SSD e a interação com outros componentes do sistema. Essas variações são refletidas nos dados coletados ao longo do período de observação, mas mantêm-se próximas à média geral de desempenho esperado para o SSD em questão. Assim, o boxplot fornece uma visão clara da dispersão dos dados, destacando a mediana e os quartis que indicam a variação dentro da amostra. A ausência de outliers extremos sugere uma consistência na performance do SSD em relação à velocidade de transferência, corroborando a estabilidade operacional do dispositivo durante o período analisado.

A Figura 2 apresenta um histograma uniforme, o qual indica uma distribuição não normal dos dados. Ao contrário de uma distribuição gaussiana, que deveria exibir uma forma semelhante à um sino simétrico em torno da média, um histograma uniforme demonstra que os valores estão igualmente distribuídos ao longo do intervalo considerado. Esse padrão uniforme sugere que não há uma concentração de dados em torno de um valor central, seguido por uma diminuição gradual em direção aos extremos, como seria típico em uma distribuição normal. Tal visualização é essencial para reconhecer a natureza não paramétrica dos dados analisados, fornecendo informações fundamentais para a aplicação de métodos estatísticos apropriados à sua distribuição.

G. Teste de hipótese I - Distribuição normal

Para verificar a normalidade dos dados, começamos utilizando o módulo *normal_test* e a função *qq_plot* da biblioteca Scipy. O módulo *normal_test* é uma ferramenta estatística que ajuda a avaliar se os dados seguem uma distribuição normal, enquanto a função *qq_plot* é utilizada para gerar o gráfico Q-Q, que é uma representação gráfica que compara a distribuição dos dados com uma distribuição normal teórica demonstrado no gráfico Q-Q.



A Figura 4 ilustra a comparação entre os quantis teóricos de uma distribuição normal e os quantis observados a partir do conjunto de dados do estudo. No gráfico Q-Q, a linha de referência de 45° representa a expectativa teórica de uma distribuição normal, onde os dados devem se alinhar de forma linear sobre essa linha para que a distribuição dos dados seja considerada normal. No entanto, ao analisar o gráfico, é evidente que há uma considerável inconsistência nos dados, pois muitos dos pontos se afastam da linha de 45°, formando uma curva nas extremidades. Esta observação sugere que a distribuição dos dados não segue uma distribuição normal, já que uma grande quantidade dos dados está claramente fora da curva esperada para uma distribuição normal.

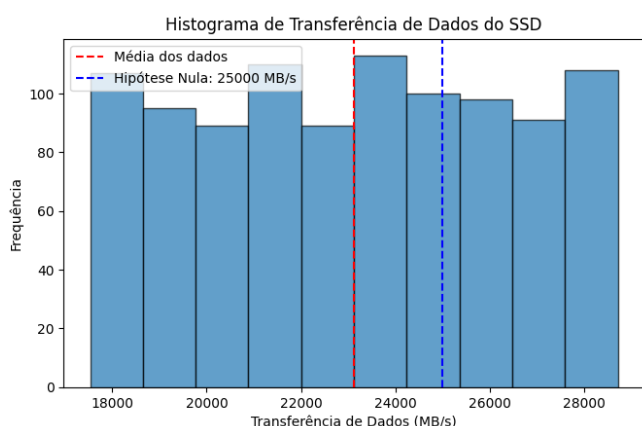
Além disso, para avaliar formalmente a normalidade dos dados, foi calculado o valor-p com base na hipótese nula (H_0) de que os dados seguem uma distribuição normal, em contraste com a hipótese alternativa (H_a) de que os dados não seguem uma distribuição normal. Usando um intervalo de confiança de 95%, o valor-p obtido foi de 0, o que é significativamente menor que o nível de significância alfa (0,05). Com isso, rejeitamos a hipótese nula e aceitamos a hipótese alternativa, confirmando que a distribuição dos dados não é normal.

H. Teste de hipótese II - Eficiência do SSD

Para determinar se o desempenho do SSD está de acordo com a qualidade prometida pelo fabricante, estabelecemos duas hipóteses estatísticas. A Hipótese Nula (H_0) assume que a média da velocidade do SSD é igual a 25.000 MB/s ($\mu = 25.000$), o valor que o fabricante garante. A Hipótese Alternativa (H_a) sugere que a média real da velocidade do SSD é diferente desse valor prometido, indicando que há uma discrepância significativa entre o desempenho real e a expectativa estabelecida.

Para testar essas hipóteses, empregamos o Teste U de Mann-Whitney, um teste estatístico não paramétrico apropriado para comparar duas amostras independentes quando não podemos assumir a normalidade dos dados. Esse mesmo teste, também chamado de Teste de Wilcoxon para amostras independentes, avalia se há uma diferença significativa entre as distribuições das duas amostras ao comparar suas medianas, sem a necessidade de pressupor uma distribuição normal dos dados.

A implementação desse teste foi realizada utilizando a função *wilcoxon* da biblioteca Scipy, uma ferramenta robusta para realizar análises estatísticas e calcular o valor-p correspondente ao Teste U de Mann-Whitney. A função *wilcoxon* é particularmente útil para esse contexto, pois adaptamos sua aplicação para comparar as amostras e determinar se a média do SSD é significativamente diferente do prometido.



Analisando a figura 5, podemos notar que o histograma representa a frequência dos dados ao mesmo tempo que nos indica onde está localizada a média dos nossos dados coletados e onde estaria localizada a média ideal. Porém, elas diferem significativamente (1867.84), um fato que resulta automaticamente na recusa da nossa hipótese nula, aceitando a hipótese alternativa. Além disso, somado com a análise gráfica, podemos chegar à conclusão de que o desempenho do SSD é muito abaixo do esperado.

IV. Conclusões

Com base nos dados coletados e nos testes de hipótese realizados, concluímos que o novo SSD não conseguiu alcançar sua meta prometida de desempenho. Os números observados mostraram uma média de velocidade significativamente inferior aos 29.000 MB/s esperados, o que indica que o produto ainda não está pronto para uso comercial ou empresarial.

Em suma, apesar das expectativas iniciais e das promessas de desempenho revolucionário, os testes realizados demonstraram que o novo SSD não está atualmente adequado para as demandas intensivas do mercado de desenvolvimento 3D. Mais refinamentos são necessários para atender às exigências rigorosas de transferência de dados e garantir a eficiência operacional esperada pelas empresas do setor.

Este relatório não apenas destaca a importância crítica da transferência de dados rápidos para o mercado empresarial, mas também ressalta a necessidade contínua de inovação tecnológica para suportar as crescentes demandas do setor de desenvolvimento 3D em um mercado global cada vez mais conectado e exigente.

V. Referências