

PERGUNTA E RESPOSTA AO PEDIDO DE ESCLARECIMENTO DA EMPRESA INTEL

O Termo de Referência prevê a adoção do “*PerformanceTest 10 CPU Mark*” (“CPU Mark”) para mensuração do desempenho de microcomputadores e notebooks a serem adquiridos onde foi estabelecida, para todos os itens, performance mínima com variação aceitável de +/- 1%. Porém, com todo o respeito, a escolha do CPU Mark como critério de mensuração de performance dos equipamentos cujos preços serão registrados não é esclarecida, ademais, não constam do Termo de Referência as razões pelas quais teriam sido fixadas as referidas pontuações de performance como critério de aceitabilidade dos equipamentos.

A NECESSÁRIA ADOÇÃO DE BENCHMARK QUE MEÇA O DESEMPENHO REAL DOS EQUIPAMENTOS

Os editais para aquisição de produtos de informática trazem nos respectivos termos de referência os requisitos técnicos mínimos dos equipamentos (quantidade de memória, capacidade de armazenamento de dados, etc.), delimitando ainda o índice (benchmark) através do qual será mensurado o desempenho do computador.

Benchmarks são programas (softwares) utilizados especificamente para avaliar de forma eficiente o desempenho de computadores. O programa tipicamente executa aplicativos e uma série de testes a partir da simulação de cargas de trabalho, gerando uma pontuação final de desempenho.

Ocorre que, tão importante quanto adotar um índice de desempenho como critério de aceitabilidade de propostas, é a escolha de um benchmark capaz de medir o desempenho real do equipamento, considerando o uso pretendido pelo órgão contratante.

A adoção de benchmark é a alternativa conhecida na indústria de tecnologia da informação para viabilizar a mensuração do desempenho real de processadores e, portanto, permitir que licitações sejam pautadas em critérios objetivos.

Como há diversos benchmarks disponíveis no mercado, a escolha do benchmark deve ocorrer em consonância com a finalidade da contratação pública. A delimitação de benchmark inadequado poderia resultar na compra de computadores com desempenho insuficiente para os fins pretendidos, o que seria contra a economicidade e o princípio da eficiência administrativa.

Há diversos índices disponíveis no mercado de informática. O CPU Mark é publicado pela Passmark Software, uma empresa com fins lucrativos^[1].

O referido benchmark testa o desempenho do processador usando um conjunto de cargas de trabalho de componentes sintéticos: Matemática Inteira, Matemática de Ponto Flutuante, Números Primos, Instruções Estendidas (SSE), Compressão, Criptografia, Física, Classificação e CPU “Single Threaded”. Cada teste sintético produz métricas individuais que resultam em uma pontuação geral de “Marca da CPU”.

Ocorre que benchmarks sintéticos como o CPU Mark são geralmente utilizados para mensurar o desempenho isolado de certos componentes, como CPU, memória ou placa gráfica. Dessa forma, os índices em questão não se prestam à mensuração do desempenho como um todo, mas tão somente de componentes isolados.

No caso em tela, é inadequada a escolha do benchmark sintético CPU Mark, uma vez que o Pregão tem por objeto a aquisição de microcomputadores e de notebooks “inteiros” e não de componentes isolados.

O CPU Mark não é utilizado para mensurar o desempenho dos computadores em situações de uso real. Para que tal finalidade fosse atingida, seria necessária a escolha de benchmark que levasse em consideração a performance do equipamento como um todo, em especial no desempenho de aplicações mais corriqueiras em ambiente profissional (editores de texto, gerenciadores de correio eletrônico, etc.).

Benchmarks são programas de computador especializados executados nos sistemas sob avaliação. O programa de benchmarking executa uma série de testes e tentativas padrão simulando cargas de trabalho no sistema e gera uma pontuação de desempenho final. A pontuação de desempenho fornece um instantâneo do desempenho do sistema nas cargas de trabalho medidas, o que permite uma comparação objetiva e baseada em dados.

Avaliar produtos usando benchmarks baseados em desempenho, em vez de números de processador ou velocidades de clock, pode levar a decisões mais bem informadas. Os benchmarks baseados em aplicativos baseados em modelos de uso podem fornecer uma estrutura sólida para comparar o desempenho de produtos de computação a serem implantados em agências governamentais.

O Termo de referencia traz ainda um problema ao utilizar-se dos resultados públicos disponíveis no site <https://www.cpubenchmark.net/>, pois os números exibidos ali são um compilado de diversos resultados submetidos que não representam as especificações técnicas do sistemas a serem adquiridos além de não serem resultados garantidos pela Passmark no que diz respeito a sua exatidão, atualidade, exaustividade e relevância para os propósitos da PRODÉRJ, deixando a responsabilidade de validação dos números por conta do interessado.

<https://www.passmark.com/legal/disclaimer.php>

Disclaimer and liability

Disclaimer

1. While the information and software contained in or made accessible by the Site has been formulated in good faith and with all due care, the Host does NOT warrant:
2. its accuracy, currency, reliability and correctness;
3. that it is free from errors or omission, or that it is definitive;

4. that it is free from computer viruses or other defects; or
5. its fitness of purpose for any use.
6. Before relying on the information contained in or made accessible by the Site, you should independently verify its accuracy, currency, completeness and relevance for your purposes, and you should obtain any appropriate professional advice.

O próprio site reconhece o problema, de que os resultados são uma média de valores e que variam com o tempo, e orientam a não utilizar diretamente os números da página publica em processos de compras.

https://www.cpubenchmark.net/graph_notes.html

1. Results are averages and vary over time

While we do perform some in house benchmarking, nearly all of the results in the graphs are averages of data sent to us by users of the PerformanceTest software. Approximately 500 new benchmark results arrive daily and each night we incorporate all the new results into the graphs. This means that the CPUMark values for each CPU model will vary over time. Popular older models have very stable scores but newer models can vary a lot, especially during the first couple of weeks following their release on the market. If you are planning on using the results as part of a tender, then either ask your suppliers to run the benchmark software on the exact hardware you are getting or take a print-out of the graphs on the day the tender is released (to avoid the values changing over time).

Ao se analisar de maneira mais profunda os resultados publicados pelo Passmark, notamos ainda que ele informa somente os 5 últimos resultados e que eles já demonstram uma variação de 3621 (27.5%) pontos em menos de 2 dias. Variação essa muito acima da aceita pelo termo de referência. Não se tem informações suficientes no site como para auditar os resultados ou qualificar as razões de tal variação.

[PassMark - Intel Core i7-10700 @ 2.90GHz - Price performance comparison \(cpubenchmark.net\)](#)

Qual benchmark usar?

Existem vários benchmarks de desempenho disponíveis, e nem sempre é fácil escolher o(s) benchmark mais apropriado(s) para uma licitação específica.

A escolha de um benchmark inadequado pode resultar na compra de um sistema de computador diferente do que a organização exige e, em alguns casos, pode levar a subestimar erroneamente os benefícios de usar fornecedores específicos e seus produtos.

Independentemente do benchmark escolhido, é extremamente importante configurar e seguir uma metodologia rigorosa ao usar benchmarks de desempenho. Variações na forma como um benchmark é executado pode levar a resultados serem pouco confiáveis e não comparáveis, o que pode até resultar em um desafio para a adjudicação do contrato.

O primeiro aspecto a considerar é o tipo de benchmark. Existem dois grupos atualmente disponíveis no mercado:

- **Os benchmarks de componentes / sintéticos** são normalmente projetados para caracterizar o desempenho isolado de certos recursos de certos componentes de hardware, como CPU, memória ou placa gráfica.
- **Benchmarks baseados em aplicação/sistema** medem casos de uso bem definidos, inclusive usando aplicações reduzidas de softwares normalmente executados. Todas as ações do usuário são definidas para que sejam medidas, incluindo tempo de processamento, erros e multitasking, de modo a avaliar o desempenho geral de um sistema

Para qualquer um desses tipos, um bom benchmark de desempenho deve sempre **ter** os atributos mínimos descritos abaixo:

- **Relevante e representativo:** As autoridades de compras governamentais devem escolher um benchmark ou uma combinação de benchmarks que medem o desempenho usando testes representativos do uso diário real para o qual o sistema se destina.

Se a referência não for relevante, as autoridades de contratação podem correr o risco de adquirir um produto diferente do que é necessário.

- **Uso de tecnologias “Future- Proof ”:** As novas compras de TI geralmente são voltadas para melhor desempenho, tecnologia mais recente e compatibilidade futura. No entanto, prever o futuro é uma coisa difícil de fazer, e as compras de TI do governo devem se concentrar no que está disponível no mercado **hoje**. Tecnologias inovadoras, como OpenCL, têm sido a “onda do futuro” há mais de 10 anos. A realidade é que não é uma tecnologia convencional **hoje**. Portanto, se não houver nenhum plano para atualizar o software por um longo período, a adoção de novas tecnologias como o OpenCL não deve ser uma preocupação. Por outro lado, adoção de Machine Learning e outras tecnologias de inteligência artificial por aplicações e softwares tem tido uma adoção crescente nos últimos anos em computação cliente, e que poderia ser uma consideração ao tentar tornar os processos de compra “Future- Proof ”.

- **Reconhecida e construída com a contribuição das partes interessadas:** As autoridades de compras devem escolher benchmarks de desempenho desenvolvidos e mantidos por consórcios da indústria, reconhecidos, ou seja, organismos de padronização independentes sem fins lucrativos com ampla representação da indústria. Benchmarks de consórcios foram abertamente desenvolvidos e amplamente debatidos, têm metodologias bem definidas e, portanto, são geralmente objetivos, imparciais, confiáveis, reproduzíveis e amplamente aceitos .

100% dos recursos de cada componente podem ser enganosos. Benchmarks sintéticos que medem as máximas capacidades de cada componente pode ser úteis para desenvolvedores de software para melhorar ou aperfeiçoar novas versões de software,

mas eles podem criar uma expectativa de desempenho que aplicativos de software padrão, **hoje**, não são capazes de alcançar . Softwares de **hoje**, com algumas exceções, não são capazes de escalar a todos os core/threads disponíveis. É essencial verificar o software em uso e os recursos que ele suporta antes de selecionar o benchmark .

- **Regras básicas de projetos de computador e comparações**

(Computer Architecture: A Quantitative Approach Fifth Edition, Hennessy & Patterson)

- **Medir o desempenho e os usos principais**

“Nossa posição é que a única medida consistente e confiável de desempenho é o tempo de execução de programas reais , e que todas as alternativas propostas para o tempo como a métrica ou para programas reais como os itens medidos eventualmente levaram a alegações enganosas ou mesmo a erros no projeto de computador.”

- **Relatório de desempenho**

“O princípio orientador do relatório de medições de desempenho deve ser reprodutibilidade - liste tudo o que outro experimentador precisaria para duplicar os resultados.”

Os benchmarks desenvolvidos de forma privada podem **ter** vários níveis de envolvimento das partes interessadas no processo de desenvolvimento. A entrada é solicitada, mas o desenvolvedor toma a decisão sobre as medições de benchmark. Além disso, normalmente há menos transparência na metodologia utilizada nesses benchmarks, o que poderia torná-los inadequados para comparar o desempenho de diferentes sistemas de informática , especialmente no contexto de licitações públicas, pois variações na metodologia influenciam nas pontuações e tornam o desempenho resultante pontuações não comparáveis.

- **Atualizados:** as autoridades de compras devem sempre usar a versão mais recente disponível de qualquer benchmark.

Bons benchmarks de desempenho são atualizados continuamente e novos benchmarks são introduzidos regularmente para acompanhar o desenvolvimento e a inovação na indústria de computadores. Um benchmark que não está atualizado não levará em consideração como os novos recursos (por exemplo, multithreading) afetam o desempenho.

Usar um benchmark desatualizado para comparar dois sistemas pode fornecer uma comparação de desempenho imprecisa. Por exemplo, suponha que um sistema ofereça benefícios reais de desempenho com tecnologia aprimorada e um segundo sistema mais antigo não. Se o benchmark desatualizado não reconhecer e medir a melhoria da tecnologia, o sistema com a tecnologia antiga pode realmente **ter** uma pontuação mais alta do que o sistema com a nova tecnologia de melhor desempenho. Em suma,

benchmarks desatualizados podem prejudicar os produtos mais recentes, inovadores e de melhor desempenho.

Sobre desenvolvedores de benchmark

Os desenvolvedores de referência são categorizados em ordem de transparência e abertura, conforme abaixo:

- Consórcio de benchmarking sem fins lucrativos (exemplos são SPEC , BAPCo e EEMBC)
- Comunidade de código aberto sem fins lucrativos (um exemplo é a Principled Technologies)
- Fornecedor independente de benchmarks com fins lucrativos (um exemplo é UL-Benchmarks)
- Desenvolvedores menores com fins lucrativos (um exemplo é Primate Labs , PASSMARK)

O processo de desenvolvimento de benchmark é diferente para cada tipo de organização. Por exemplo, cada membro de um consórcio de benchmarking é uma parte interessada e pode ajudar na codificação, oferecer feedback para o desenvolvimento e/ou participar de testes de pré-lançamento. Cada membro de um consórcio de benchmarking também tem um voto em relação aos recursos que serão incluídos no benchmark final lançado. A Intel faz parte desses consórcios como contribuinte para o desenvolvimento desses benchmarks como qualquer outro membro parceiro. Exemplos de tais consórcios incluem SPEC , BAPCo e EEMBC.

Links do site:

SPEC <http://www.spec.org>

BAPCo <http://www.bapco.com>

EEMBC <http://www.eembc.org>

No caso de uma comunidade de código aberto ou fornecedor independente de benchmark, as decisões finais sobre quais recursos serão incluídos em um benchmark final lançado são normalmente tomadas pelos executivos da organização. Os fornecedores independentes de benchmarks variam no nível de entrada do usuário que procuram. Por exemplo, a entrada de UL-Benchmarks solicita de sua comunidade de usuários, embora UL -Benchmarks executivos a tomar as decisões finais.

As pequenas empresas tendem a ser as menos transparentes de todos os desenvolvedores de benchmark, porque normalmente não têm recursos para solicitar feedback do usuário durante o processo de desenvolvimento de benchmark.

Intel apoia clientes em todo o mundo com a melhor informação disponível . Estamos envolvidos no desenvolvimento de benchmarks (SPEC e BAPCo) e em sua promoção como melhor prática para comparação de sistemas computacionais, um exemplo é a recomendação de aquisição neutra de produtos para os governos da BITKOM (Alemanha) <https://www.itk-beschaffung.de/EN>. Em colaboração com a indústria líderes como HP *, Lenovo *, ACER *, Dell * e AMD * e representantes do Governo alemão, o bitkom vem desenvolvendo e atualizando suas recomendações para aquisições de equipamentos de TI desde 2008 . Em 2020, a UE adotou esse modelo e

you can find the information in https://ec.europa.eu/info/policies/public-procurement/tools-public-buyers/sector-specific-tools_pt

Based on this, we would like to request that the reference term be reviewed to include a relevant, representative and reproducible benchmark.

RESPOSTA: Entendemos que o benchmark atualmente especificado no termo de referência é ajustado às necessidades técnicas atuais. No entanto para as próximas contratações estaremos avaliando as opções sugeridas.