

Enfrentando os desafios das cargas de trabalho de IA com o portfólio de IA da Dell

Uma comparação do portfólio de IA da Dell com ofertas semelhantes da HPE

A adoção da IA apresenta novos desafios para as empresas e para sua equipe de TI de data center, incluindo:

- Lidar com a falta de habilidades da equipe atual por meio de treinamento ou contratação externa
- Compreender as necessidades de preparação de dados da IA, incluindo a qualidade, quantidade, localização e estado atual dos dados dos negócios
- Avaliar as metas dos negócios para determinar quais modelos e implementações de IA apresentarão benefícios
- Avaliar as necessidades computacionais, de rede e de armazenamento dos sistemas planejados de IA e adquirir esses sistemas

Com portfólios prontos para IA, os fornecedores de infraestrutura, como a Dell Technologies e a HPE, oferecem soluções integradas que abrangem todo o ciclo de vida da IA. Pesquisamos informações publicamente disponíveis sobre os portfólios de IA da Dell e da HPE e examinamos testes de referência de desempenho do MLPerf®. Descobrimos que a Dell está bem posicionada para ajudar as empresas a adotar a Inteligência Artificial com um portfólio que compreende opções de servidor e armazenamento de alto desempenho, soluções validadas e serviços profissionais que guiam o processo desde o planejamento até a produção.

Referência de desempenho do modelo de IA: comparação de resultados do MLPerf

Os testes de referência de desempenho do MLPerf® disponíveis publicamente mostram que as ofertas do portfólio de IA da Dell entregam desempenho sólido e consistente para cargas de trabalho de IA. O MLPerf® testa o desempenho tanto para treinamento quanto para inferência em vários modelos de IA. Os dados mencionados neste relatório são baseados nos resultados do MLPerf® v3.1 Inference Datacenter publicados no site da MLCommons® em novembro de 2023.¹ Comparamos ambos servidores de 8 GPUs e 4 GPUs e incluímos um resultado de cada categoria neste resumo; para ver todos eles, leia o relatório completo [aqui](#).

*Com base em uma análise da Dell, agosto de 2023. A Dell Technologies oferece soluções projetadas para oferecer suporte a cargas de trabalho com IA desde PCs/workstations (móveis e fixas) até serviços para computação com alto desempenho, armazenamento de dados, infraestrutura definida por software nativa na nuvem, switches de rede, proteção de dados, HCI e serviços. Dados de: "Robert McNeal, 'Dell, VMware and NVIDIA Bring AI to Your Data'", acessado em 17 de janeiro de 2024, <https://www.dell.com/en-us/blog/dell-vmware-and-nvidia-bring-ai-to-your-data/>.

1. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results", acessado em 12 de dezembro de 2023, <https://mlcommons.org/en/inference-datacenter-31/>.

Nossa pesquisa descobriu que a Dell tem:



O mais amplo portfólio de IA generativa*



De acordo com os resultados do MLPerf®, o servidor PowerEdge XE8640 com quatro GPUs NVIDIA H100 SXM5 atingiu a maior taxa de transferência de IA entre todos os envios de quatro GPUs, em nove categorias diferentes



Um portfólio mais amplo de serviços profissionais personalizados para IA

Comparação do desempenho do MLPerf para servidores de oito GPUs

Nos resultados do MLPerf® v3.1 para servidores com oito GPUs, o Dell PowerEdge XE9680 com GPUs NVIDIA H100 SXM5 superou em cerca de 4,25x o HPE ProLiant XL675d Gen10 Plus com GPUs NVIDIA A100 SXM4 (consulte a Figura 1).

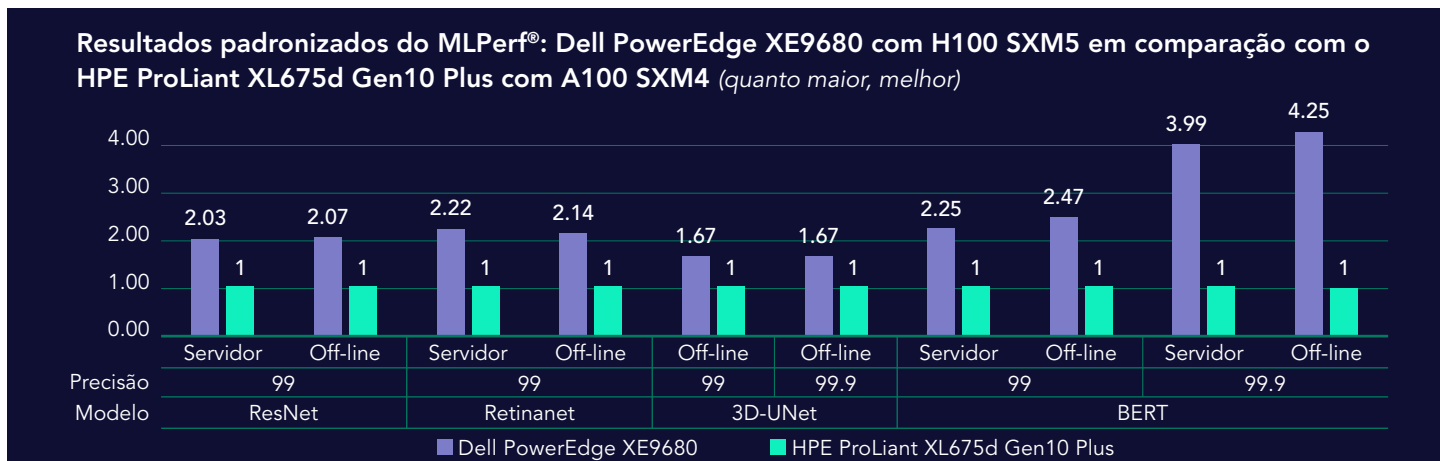


Figura 1: Resultados do MLPerf® publicados para o Dell PowerEdge XE9680 e o HPE ProLiant XL675d Gen10 Plus em 29/11/23. O sistema Dell usa GPUs NVIDIA H100, enquanto o sistema HPE usa GPUs da geração anterior. Fonte: Principled Technologies usando os dados da MLCommons®.*

Comparação do desempenho do MLPerf para servidores de quatro GPUs

O servidor PowerEdge XE8640 com quatro GPUs NVIDIA H100 SXM5 atingiu a maior taxa de transferência de IA entre todos os envios de quatro GPUs, em nove categorias diferentes. Conforme mostrado na Figura 2, em comparação com o servidor HPE ProLiant DL380a, o servidor Dell atingiu uma pontuação até 2,07 vezes maior de acordo com a referência de desempenho MLPerf®.

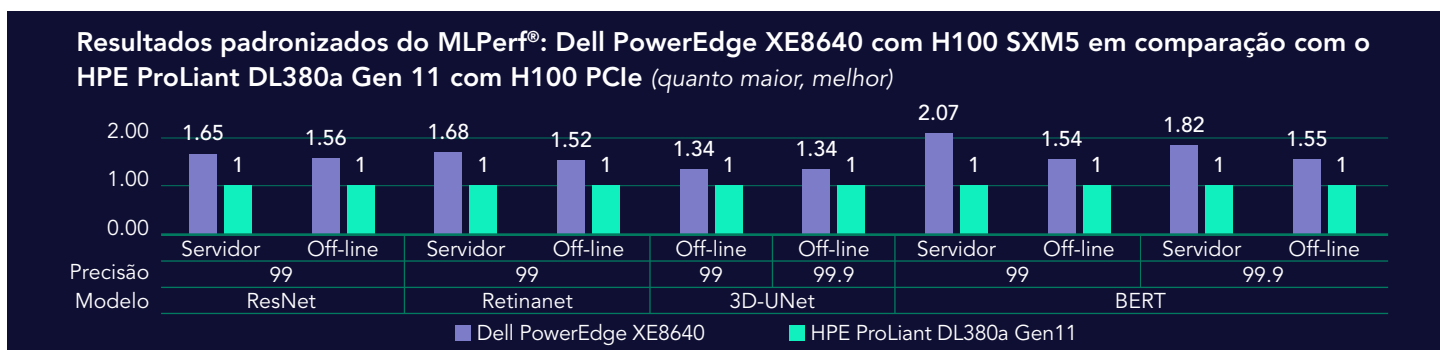


Figura 2: Resultados do MLPerf® publicados para o Dell PowerEdge XE8640 e o HPE ProLiant DL380a Gen11 em 29/11/23. O sistema Dell usa o formato NVIDIA H100 SXM, enquanto o sistema HPE usa o formato PCIe, menos potente. Fonte: Principled Technologies usando os dados da MLCommons®.*

Em resumo: Contribuidores adicionais para o sucesso da IA

Os modelos de IA exigem mais do que servidores de alto desempenho para o sucesso. Você também deve considerar o armazenamento para dados não estruturados e serviços profissionais para planejar, preparar, implementar e gerenciar sua solução de IA. O portfólio da Dell oferece armazenamento para conjuntos de dados de IA com a série PowerScale para armazenamento de arquivos e armazenamento Elastic Cloud Storage ou ObjectScale para armazenamento em objeto.

As organizações podem aproveitar as vantagens dos serviços profissionais e de consultoria da Dell para IA, que oferecem alguns serviços que a HPE não oferece, como a preparação de dados. O portfólio da Dell também inclui designs validados para IA, o que elimina as suposições no projeto e na implementação da IA.

[Leia o relatório](#) ▶

*Pontuação encerrada da inferência v3.1 verificada pelo MLPerf. Recuperado de <https://mlcommons.org/benchmarks/inference-datacenter/> em 5 de dezembro de 2023, entradas 3.1-0069, 3.1-0085, 3.1-0067 e 3.1-0084. O nome e logotipo MLPerf são marcas registradas e não registradas da MLCommons. Association nos Estados Unidos e em outros países. Todos os direitos reservados. Uso não autorizado estritamente proibido. Acesse www.mlcommons.org para obter mais informações.

▶ Consulte a versão original deste relatório, em inglês, em <https://facts.pt/PIsGyY>



Facts matter.®

Principled Technologies é uma marca registrada da Principled Technologies, Inc. Todos os outros nomes de produtos são marcas comerciais de seus respectivos proprietários. Para obter mais informações, consulte o relatório.