

Você está pronto para a próxima onda de IA no seu data center?

Não há motivo para adivinhar. Descubra o que profissionais como você estão fazendo agora mesmo e como se preparar para o que vem a seguir.



Smarter
technology
for all

Lenovo

Não tenha medo. Esteja preparado.

As ferramentas de IA estão aumentando as demandas sobre os servidores dos data centers e continuarão fazendo isso no futuro previsível. Você precisa de um plano para lidar com esses workloads crescentes. A boa notícia é que há muita ajuda disponível. Na verdade, você pode até se surpreender com as oportunidades que estão à sua espera.

Neste guia, você vai explorar:

O estado da IA nos data centers	3
Estratégias para acelerar a AI	6
Como integrar tudo	8
Como tornar tudo mais eficiente	13



Passe o cursor sobre os círculos vermelhos ao longo do guia para obter mais informações.





Os data centers estão adotando IA, estejam prontos ou não

Os líderes de TI querem colocar a IA para trabalhar o quanto antes.



O uso de tecnologias emergentes como Gen AI foi a prioridade de negócio número 1 neste ano.¹



dos CIOs esperam ganhos de produtividade derivados da IA nos próximos 1 a 2 anos.²



dos CIOs dizem que usar IA facilitará o cumprimento das metas de sustentabilidade de TI de sua organização.²

Por que a adoção de IA está crescendo? Porque está funcionando.



O data center médio não está pronto

Muitos data centers já estão operando no limite ou perto do limite em espaço, energia ou ambos⁷. Essas demandas também alcançam as equipes que administram os dados e 84% dizem que seu workload está acima da capacidade.⁸

Tanto as pessoas quanto os recursos estão no limite. Para muitas organizações, adicionar mais profissionais de suporte pode não ser possível. 45% dos CIOs dizem que é difícil encontrar talentos relacionados à IA para suas equipes.¹



A IA não vai esperar os data centers estarem prontos

Os líderes das organizações enxergam as oportunidades da IA e não estão esperando para investir.

45% dos executivos dizem que já aumentaram seus investimentos em IA.⁹

96% dos CIOs antecipam um aumento no investimento em IA.¹⁰

79% dos líderes dizem que sua organização precisa adotar IA para continuar competitiva.¹⁰

A inferência coloca a IA ao alcance

Os conjuntos de dados usados para treinar modelos fundacionais cresceram em uma escala que exige grandes quantidades de hardware de computação especializado, tradicionalmente reservado a organizações acadêmicas, financeiras e governamentais.

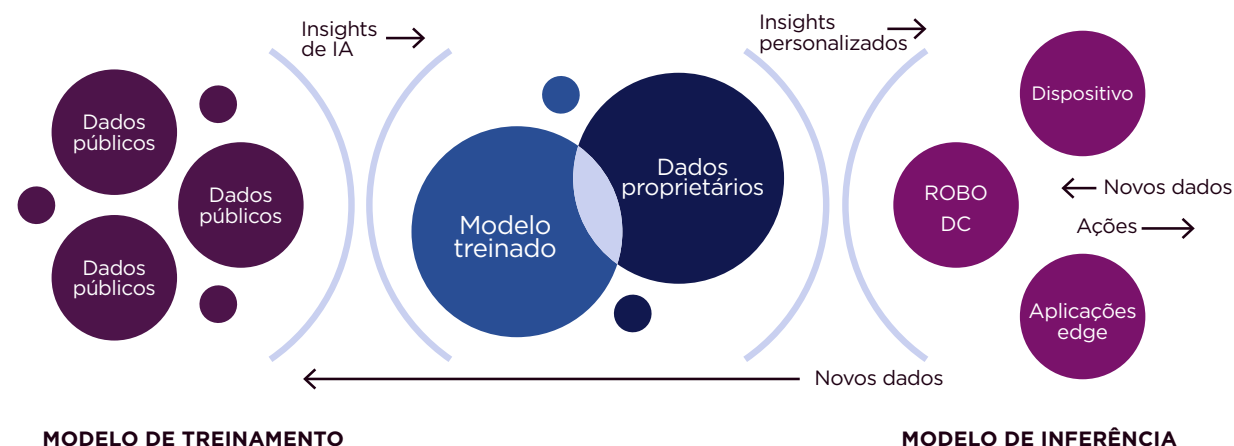
A inferência de IA pega os modelos já treinados e os aplica a novos conjuntos de dados proprietários para tarefas específicas de aplicação. Graças às menores exigências de processamento, a inferência de IA está abrindo as portas para que mais empresas e organizações de todos os portes aproveitem a IA.

A IA está se democratizando por meio da inferência. A disposição para investir está crescendo. Então, o que está impedindo isso de avançar?



A quantidade de poder computacional necessária para treinar os maiores modelos de IA está dobrando a cada 3 a 10 meses.¹¹

Systems of action





Investir não é tudo

Embora a IA já esteja ao alcance e pareça haver orçamento disponível, sem uma implementação adequada esse investimento pode gerar aumento no custo total de propriedade, ou TCO, devido a workloads adicionais, maiores demandas de energia e ameaças à segurança.

Por exemplo, embora apenas 39% dos CIOs tenham dito que sua tecnologia de segurança estava pronta para IA, 72% dos CEOs afirmam que estão agindo em IA sem uma avaliação dos impactos potenciais. Os colaboradores sabem disso.^{2,12}



dos trabalhadores dizem que a liderança de sua organização não tem um plano para implementar IA.¹⁰

Um plano de IA também deve incluir uma estratégia de eficiência energética, que pode ser o aspecto de maior impacto do crescimento da IA no data center. A eletricidade representa 46% do gasto total, o que a torna o maior custo contínuo para os data centers empresariais¹³. A crescente demanda por workloads de IA deve provocar um aumento significativo na capacidade, no consumo de energia e nas emissões de carbono dos data centers.

Avançar sem um plano para preparar seu data center para essas mudanças pode gerar mais atrasos, ineficiências, gastos e brechas de segurança. Essa pode ser a razão pela qual os CIOs citam a velocidade como sua barreira número 1 para adotar IA.²

Obter o máximo valor da IA exige compreender seus casos de uso e determinar quais ferramentas podem escalar junto com sua organização à medida que ela cresce.



6 princípios para acelerar as experiências de IA no seu data center

As ferramentas de IA provavelmente continuarão melhorando. Para acompanhar esse ritmo, você vai precisar de uma nova estratégia, que reconheça as demandas sobre as equipes de TI e os orçamentos, ao mesmo tempo em que permita adaptabilidade e escalabilidade conforme essas demandas evoluem.

1

Leve em conta seu TCO

À medida que as workloads de IA escalam, tudo escala. De hardware adicional a pessoal adicional, o custo total de propriedade da IA deve ser claramente compreendido. Se sua organização precisa cumprir regulamentações de privacidade de dados, também devem ser considerados custos adicionais de segurança e estratégias de mitigação de perdas. À medida que o acesso à IA se democratizou, ficou mais fácil começar sem saber o quão sustentável seu modelo pode ser. Uma avaliação aprofundada dos seus recursos e necessidades poderá dar uma visão muito mais clara do seu TCO.

2

O processamento importa

Suas necessidades de inferência de IA importam na hora de encontrar a plataforma certa, equilibrando potência e eficiência. Os CPUs mais recentes estão sendo projetados para IA, são mais eficientes e até mais eficazes no desenvolvimento, teste e treinamento de modelos pequenos e médios. Ao desenvolver e treinar modelos de IA maiores, os GPUs podem ser mais rápidos e reduzir o custo total de propriedade. Para obter o melhor dos dois mundos, adote uma plataforma flexível o suficiente para oferecer todos os benefícios dos CPUs e com capacidade de escalar com eficiência para o poder de GPU quando necessário.

3

Otimize para a eficiência

Minimize seu impacto financeiro e ambiental enquanto maximiza o desempenho. Consolide o número de servidores físicos necessários, implemente sistemas modernos de refrigeração e use ferramentas de administração projetadas especificamente para sua tecnologia de servidor para gerenciar de forma preditiva a energia e a localização das workloads. Certifique-se de que seus parceiros estejam considerando o impacto ambiental em tudo, da embalagem às compensações de CO₂.



4

Torne a confiabilidade algo empolgante

Às vezes, a melhor tecnologia se torna quase invisível. Relatórios de confiabilidade de servidores estão disponíveis há anos, portanto nunca há motivo para se perguntar se o seu vai resistir. O mais importante na era da IA é saber se o seu servidor é inteligente o suficiente para prever quando uma manutenção poderá ser necessária ou onde uma nova eficiência pode ser encontrada. É aí que um hardware de servidor confiável, combinado com um software de gerenciamento excepcional, pode fazer toda a diferença. Sempre considere a combinação, nunca apenas o hardware isoladamente.

5

Construa sobre uma base segura

Desempenho, eficiência e confiabilidade não significam nada se seus sistemas não estiverem protegidos. A IA pode dominar as manchetes, mas entre os CIOs, a cibersegurança está empatada com a IA como a prioridade mais urgente². Os data centers devem buscar soluções de IA com camadas de segurança que vão da cadeia de suprimentos do servidor até a sofisticação do software e até dos serviços. Um parceiro confiável é tudo quando se trata de proteger um data center pronto para IA.

6

Foque na inovação contínua

Existe uma razão financeira sólida para incentivar a inovação contínua. Empresas com culturas fortes de inovação têm menos chance de encontrar barreiras em suas transformações e muito mais chance de liderar seu setor no desenvolvimento de produtos e no atendimento à demanda emergente de clientes por meio de operações impulsionadas por IA.¹⁴ Mantenha-se atualizado com os avanços em tecnologia de IA e em práticas de gerenciamento de data centers. Atualize regularmente hardware e software e faça parceria com um serviço gerenciado confiável para assumir as tarefas rotineiras, de modo que sua equipe tenha mais liberdade para iniciativas inovadoras.



“Smarter” está impulsionando o futuro da IA

Os servidores Lenovo ThinkSystem® impulsionados por processadores AMD EPYC™, combinados com aceleradores AMD Instinct™, oferecem uma solução poderosa e com bom custo-benefício on premises ou em private cloud para uma ampla variedade de necessidades de aplicações e modelagem de IA.

A GPU discreta AMD Instinct™ MI300X é baseada na arquitetura de próxima geração AMD CDNA™ 3, com processamento veloz, funções específicas para IA e memória HBM3 de 192 GB sem precedentes. Criado com os aceleradores AMD Instinct em mente, o AMD ROCm™ open software ajuda a entregar soluções de alto desempenho prontas para uso em treinamento e inferência de IA, permitindo que as empresas avancem em seus objetivos de negócio.

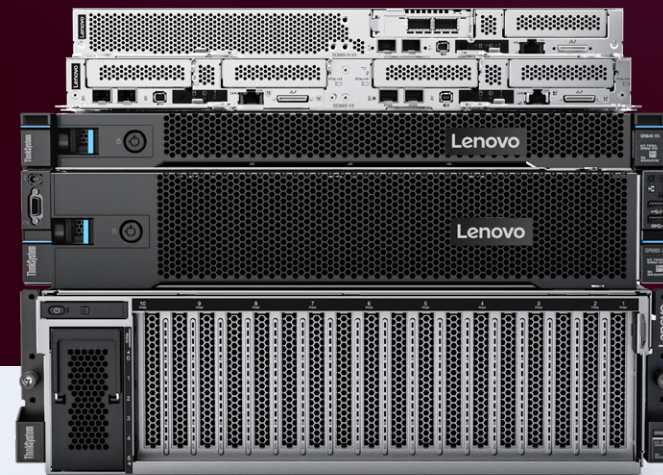
VÍDEO

Veja como a Nscale entrega IA[Ver agora >>](#)

ThinkSystem SR685a V3



Lenovo Infrastructure Solutions impulsionadas pela mais recente geração de arquitetura de servidores, os processadores AMD EPYC Series



Juntos, Lenovo e AMD oferecem aos data centers um nível excepcional de:

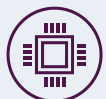
**Performance****Security features****Reliability****Efficiency**



Experimente o desempenho da próxima geração

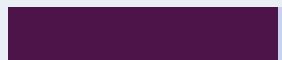
Juntos, os servidores Lenovo com processadores AMD EPYC™ estão respondendo às crescentes demandas de TI atuais com desempenho líder do setor, com 173 recordes mundiais de benchmark¹⁵.

Os servidores ThinkSystem com processadores AMD EPYC podem:



100%

suportar 100% mais núcleos do que os servidores ThinkSystem da geração anterior.¹⁶



98%

alcançar até 98% de melhoria na largura de banda de streaming de memória com DDR5 em comparação com DDR4.¹⁸



88%

alcançar até 88% melhor desempenho para aplicações de virtualização do que a geração anterior.¹⁷



O dobro

da taxa de transferência de dados com PCI Gen 5 para inferência de IA e machine learning mais rápidos.¹⁹



Impulsionado por AMD Instinct™ MI300X



Até 1,3x

o desempenho de IA em comparação com aceleradores concorrentes.²⁰

CASO DE SUCESSO

5 vezes mais rápido em simulações complexas

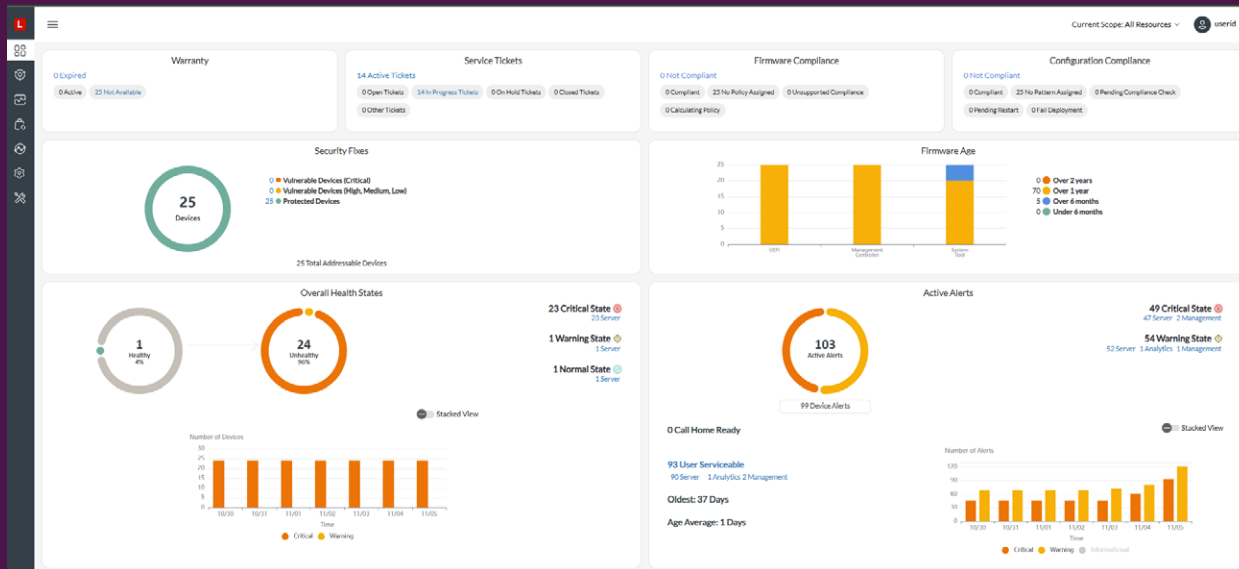
Uma consultoria conseguiu ajudar equipes de design e engenharia a executar simulações complexas até 5 vezes mais rápido com servidores Lenovo ThinkSystem equipados com processadores AMD EPYC.

[Leia mais >>](#)





INTEGRANDO TUDO



Lenovo XClarity™ Administrator

Lenovo XClarity™ Orchestrator

Lenovo XClarity™ Controller



Recupere a confiança com confiabilidade sólida como rocha

Durante 10 anos consecutivos, os servidores Lenovo foram classificados como os servidores x86 mais confiáveis do setor.²¹

O Lenovo XClarity™ One ajuda a manter essa confiabilidade no longo prazo por meio de manutenção preditiva e análise inteligente para melhorar o desempenho, a confiabilidade e a eficiência geral da sua infraestrutura.

CASO DE SUCESSO

Disponibilidade permanente para sistemas de missão crítica

Os servidores Lenovo ThinkSystem impulsionados por processadores AMD EPYC™ de alto desempenho ajudaram um porto sul-americano a oferecer sistemas de segurança confiáveis 24/7 com um aumento de 4 vezes no desempenho computacional.

[Leia mais >>](#)



Encontre eficiências de próximo nível

Os servidores Lenovo mais recentes com processadores AMD EPYC™ estão entre os servidores mais eficientes energeticamente do mundo. O desempenho líder por watt dos servidores Lenovo com processadores AMD pode ajudar a reduzir significativamente o consumo de energia.

Os servidores Lenovo V3 usam os mais recentes processadores AMD EPYC, que oferecem até 2,5 vezes a eficiência energética de outros processadores.²²

Servidores Lenovo selecionados com Neptune Direct Water Cooling permitem aumentar sua eficiência energética em até 40%²³. A tecnologia de refrigeração Lenovo Neptune pode capturar até 100% do calor do sistema, dependendo do ambiente específico, usando temperatura da água de até 45 °C.²³

Explore oportunidades para consolidar

Os processadores AMD EPYC de quarta geração oferecem consolidação de servidores de até 3 para 1 em comparação com servidores mais antigos, entregando velocidade e eficiência para tarefas de IA e ML, enquanto reduzem a pegada do data center de forma econômica.**



CASO DE SUCESSO

40% melhor desempenho com 30% menos energia e calor

Os servidores Lenovo com processadores AMD EPYC trabalharam em conjunto para ajudar um provedor líder de managed services na França a manter alto desempenho enquanto reduzia o consumo de energia em seus novos data centers.

[Leia mais >>](#)

30%
menos consumo
de energia

30%
de redução nas
emissões de calor

40%
de aumento no
desempenho das
aplicações

The background of the advertisement features a man and a woman in a server room. The man, wearing a white shirt and a lanyard, is looking at a laptop held by the woman. The woman, wearing a red cardigan, glasses, and a lanyard, is looking at the laptop screen. They are standing in front of server racks. The Lenovo logo is in the bottom right corner.



Proteja seu data center desde a cadeia de suprimentos

Os servidores Lenovo foram classificados como os servidores x86 mais seguros do setor nos últimos cinco anos.²¹

Protegendo seus dispositivos da cadeia de suprimentos ao sistema operacional e à cloud, o Lenovo ThinkShield oferece proteções em todos os níveis, incluindo IA para detecção e resposta a ameaças.

De fato, a cadeia de suprimentos da Lenovo está entre as 10 melhores do mundo, considerando todos os setores, de acordo com o Gartner.²⁴

As capacidades de segurança do Lenovo ThinkShield são fortalecidas com AMD

Infinity Guard nos servidores ThinkSystem que usam processadores AMD EPYC™.*** Projetado para ser altamente resistente aos ataques sofisticados de hoje, o AMD Infinity Guard oferece um conjunto líder de funções modernas de segurança para ajudar a proteger dados sensíveis, evitar tempo de inatividade e reduzir o desgaste de recursos.

Juntos, Lenovo e AMD ajudam a se defender contra ameaças internas e externas, diminuindo as superfícies potenciais de ataque onde dados críticos são processados.

CASO DE SUCESSO

Ganhar participação de mercado com novas capacidades

Uma empresa de sistemas de vigilância e segurança de última geração utilizou servidores Lenovo equipados com processadores AMD EPYC para habilitar capacidades líderes de mercado com sua plataforma de próxima geração.

[Leia mais >>](#)





TORNANDO TUDO MAIS EFICIENTE

Você não precisa entrar na IA sozinho

O Lenovo AI Services oferece uma solução de fornecedor único para orientar o desenvolvimento e a implementação de IA específica para cada propósito e de fácil implementação.

Lenovo AI Innovators

O ecossistema de parceiros de IA da Lenovo, chamado AI Innovators Program, oferece soluções completas de IA para acelerar o desenvolvimento e a implementação. A Lenovo avalia cada empresa antes que ela se torne parceira do AI Innovators Program, que inclui:

- ✓ Mais de 50 provedores de soluções de IA
- ✓ Mais de 165 soluções empresariais de IA
- ✓ Mais de 180 países atendidos
- ✓ Mais de 30 channel partners

Lenovo AI Discover Centers of Excellence

A Lenovo oferece acesso a data scientists, arquitetos de IA e engenheiros da Lenovo para ajudar você a avaliar soluções, realizar provas de conceito, ou POCs, e otimizar configurações.

Lenovo Responsible AI Committee

O Lenovo Responsible AI Committee garante que as soluções e os parceiros cumpram requisitos que ajudem a proteger os usuários finais e assegura que a IA seja usada de forma justa, ética e responsável. A Lenovo também assinou recentemente o AI Pact da Comissão Europeia, o Compromisso da UNESCO para uma IA Responsável e o Código Voluntário de Conduta do Canadá sobre o Desenvolvimento e a Gestão Responsável de Sistemas Avançados de IA Generativa.

A participação da Lenovo nesses acordos destaca nossa abordagem proativa para nos prepararmos para o ambiente regulatório de IA que está surgindo e transformar essas considerações em uma base para os serviços e soluções relacionados à IA.



Nós cuidamos disso para você

Lenovo TruScale™ Infrastructure as a Service (IaaS)

Cada vez mais organizações estão recorrendo a soluções as a service. De fato, 92% dos CIOs estão considerando adicionar novas ofertas aaS nos próximos dois anos.²⁵ O Lenovo TruScale™ IaaS dá acesso ao portfólio tecnológico mais amplo do setor e tira a carga de gerenciamento das equipes de TI, liberando-as para se concentrar em iniciativas mais estratégicas.



Os modelos tecnológicos flexíveis pay as you go oferecem escalabilidade de maneiras que os métodos tradicionais de aquisição não conseguem.



Libere capital e melhore a produtividade dos colaboradores.



Equipamentos, serviços e suporte podem ser adicionados, modificados ou retirados facilmente por meio de um único ponto de contato.



Monitore e avalie o estado e o uso dos dispositivos com análise proativa e preditiva.

Lenovo Professional Services

A Lenovo pode avaliar e desenhar a melhor forma de gerenciar a arquitetura de ponta a ponta, a instalação de hardware, a migração de dados e o deployment do sistema, acelerando seu tempo até a produtividade. O Lenovo Professional Services também pode oferecer experiência e cobertura global para atender lacunas de recursos.

Lenovo Managed Services

O Lenovo Managed Services ajuda você a fechar lacunas de recursos, melhorar o desempenho, limitar o tempo de inatividade e trazer mais tranquilidade ao ajudar a controlar e prever seus custos, a disponibilidade do sistema e a capacidade.

Lenovo Sustainability Support

O serviço de transporte low carbon da Lenovo entrega equipamentos de TI por via aérea usando combustível de aviação sustentável. O Lenovo Asset Recovery Services mantém os recursos em circulação reutilizando peças quando possível, recondicionando dispositivos e reciclando ativos obsoletos de forma ambientalmente responsável. O Lenovo CO₂ Offset Services facilita a compensação das emissões estimadas de carbono das suas soluções Lenovo.



Lenovo e AMD aceleram a IA para o seu data center

Descubra como Lenovo e AMD podem ajudar você a entregar as workloads de alto desempenho necessárias para a inovação com IA, ao mesmo tempo em que ajudam a melhorar a eficiência, acelerar o time to value e reduzir seu custo total de propriedade.

Leia mais >>



**Smarter
technology
for all**

Lenovo

Sources

- 1 IDC e Lenovo, "CIO Playbook 2024", 2024
- 2 Lenovo, "Inside the Tornado: Lenovo Global CIO Report 2024", abril de 2024
- 3 Accenture, "Uma nova era de IA generativa para todos", 2023
- 4 McKinsey, "O potencial econômico da IA generativa: a próxima fronteira da produtividade", junho de 2023
- 5 National Bureau of Economic Research, "O impacto potencial da inteligência artificial nos gastos com saúde", janeiro de 2023
- 6 Built In, "21 exemplos de IA em finanças", outubro de 2023
- 7 Análise baseada em dados internos da AMD
- 8 Foundry, "State of the CIO Survey 2023", janeiro de 2023
- 9 Gartner, "Gartner identifica as principais tendências que moldam o futuro da ciência de dados e do aprendizado de máquina", agosto de 2023
- 10 Microsoft e LinkedIn, "Relatório Anual do Índice de Tendências de Trabalho de 2024", maio de 2024
- 11 Accenture, "Visão Tecnológica 2023", março de 2023
- 12 IBM Institute for Business Value, "Tomada de decisão do CEO na era da IA: agir com intenção", junho de 2023
- 13 IDC, "Relatório da IDC revela crescimento impulsionado por IA no consumo de energia de data centers, prevê aumento nos gastos com instalações de data centers em meio ao aumento dos custos de eletricidade", setembro de 2024
- 14 McKinsey, "Como empresas inovadoras alavancam a tecnologia para superar o desempenho", dezembro de 2023
- 15 AMD, "Recordes Mundiais do Processador AMD EPYC™", setembro de 2024
- 16 Lenovo, "Lenovo ThinkSystem SR665 V3 Server Product Guide", setembro de 2024
- 17 resultados do VMmark® 3.1.1, em 18 de outubro de 2023. Consulte <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2021-05-04-Lenovo-ThinkSystem-SR665.pdf> para mais detalhes. VMware® e VMmark® são marcas registradas ou marcas comerciais da VMware, Inc. VMware VMmark é um produto da VMware, Inc.
- 18 Baseado em testes internos da Lenovo que comparam a largura de banda de memória dos servidores Lenovo ThinkSystem V3 com AMD Genoa, 12 canais de DDR5 @ 4800, versus servidores Lenovo ThinkSystem com AMD Milan, 8 canais de DDR4 @ 3200.
- 19 TechTarget, "PCIe SSD (unidade de estado sólido Peripheral Component Interconnect Express)," 2023
- 20 MI300-17: Medições realizadas pelo AMD Performance Labs em 11 de novembro de 2023
- 21 ITIC, "ITIC 2023 Global Server HW, Server OS Reliability Report," setembro de 2023
- 22 WSJ, "Advancing Data Center Sustainability, AMD desenvolve designs de processador inovadores para tornar a computação mais eficiente em termos energéticos", dezembro de 2024
- 23 Lenovo, "Lenovo Think System SD665 V3 Guia de produto do servidor Neptune DWC", dezembro de 2023
- 24 Lenovo, "Lenovo ocupa o 10º lugar no relatório Gartner® Supply Chain Top 25 para 2024", maio de 2024
- 25 Lenovo, "O CIO Resiliente", maio de 2023

Disclaimers

* O servidor Lenovo SR665 V3 ocupa o lugar número 1 no benchmark padrão do setor SPECpower_ssj2008 de eficiência de desempenho por watt, em 17 de outubro de 2023. 34.597 ssj_ops/watt no geral. Consulte https://spec.org/power_ssj2008/results/res2023q4/power_ssj2008-20230926-01323.html para detalhes do resultado. Consulte <http://www.spec.org> para informações adicionais. SPEC®, SPEC ACCEL®, SPEC CPU®, SPEC MPI®, SPEC OMP®, SPEC VIRT_SC®, SPEC VIRT®, SPECchpc™, SPECjbb® e SPECpower_ssj® são marcas comerciais da Standard Performance Evaluation Corporation, ou SPEC.

** Resultados de VMmark 3.1.x, em 20 de outubro de 2023. Dois servidores Lenovo ThinkSystem SR665 V3, cada um com dois processadores AMD EPYC 9654, obtiveram 40.66 @ 42 tiles. Consulte <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2023-06-13-Lenovo-ThinkSystem-SR665V3.pdf> para mais detalhes. Dois servidores HPE, cada um com dois processadores AMD EPYC 7702, obtiveram 12.78 @ 14 tiles. Consulte <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2019-08-07-HPE-ProLiant-DL385Gen10.pdf> para mais detalhes. Para saber mais sobre VMmark, visite <https://www.vmware.com/products/vmmark.html>. VMware® e VMmark® são marcas registradas ou marcas comerciais da VMware, Inc. VMware VMmark é um produto da VMware, Inc. Os resultados reais de consolidação variarão conforme muitos fatores.

*** As funções do AMD Infinity Guard variam conforme as gerações de EPYC™ Processor. As funções de segurança do Infinity Guard devem ser habilitadas pelos OEMs de servidores e ou provedores de serviços em nuvem para funcionar. Consulte seu OEM ou provedor para confirmar a compatibilidade com esses recursos. Obtenha mais informações sobre Infinity Guard em <https://www.amd.com/en/technologies/infinity-guard>.