

# A Lenovo e a NVIDIA impulsionam a Plataforma como Serviço (PaaS) de Modelos de IA da Edgebricks

**A plataforma da Edgebricks oferece uma Infraestrutura de Aprendizado de Máquina Totalmente Automatizada para construir, treinar e implantar modelos de IA 10 vezes mais rápido**

## Resumo

Atualmente, 89% das empresas buscam fazer mais com a adoção de IA e Aprendizado de Máquina em seus produtos existentes, processos e trabalho cotidiano. A IA generativa está causando um aumento adicional nas aplicações potenciais em 100 vezes, à medida que passamos de uma IA baseada em classificação para gerar conteúdo para imagens, vídeos, código, texto e outros tipos de mídia.

A construção e execução de modelos privados GPT são essenciais para muitas indústrias devido a questões de conformidade, privacidade de dados e custos. O Aprendizado de Máquina também é a workload mais cara para hospedar em nuvens públicas, uma vez que requer alta capacidade de computação, memória, largura de banda de rede e armazenamento.

A plataforma de IA/ML da Edgebricks, como serviço, ajuda as empresas a construir, treinar e implantar IA/ML 10 vezes mais rápido usando os servidores Edge da Lenovo com GPUs da NVIDIA no local de sua escolha. Assim, elas obtêm a facilidade de uso de uma nuvem pública, mas com o controle e a flexibilidade de uma nuvem privada de IA/ML.

## Desafio

O crescimento da IA Empresarial e a adoção da IA Generativa estão levando as empresas a adotarem IA/ML para permanecerem competitivas, aumentarem a produtividade de suas equipes e reduzirem custos. Elas precisam construir a infraestrutura de IA/ML por conta própria ou aproveitar soluções genéricas em nuvens públicas. No entanto, a maioria delas menciona a contratação do talento certo para IA/ML como um de seus maiores problemas.

O conjunto geral de ferramentas e o cenário estão evoluindo rapidamente, e não há uma maneira fácil de construir infraestrutura e pipelines de IA/ML sem alto custo e experiência. As nuvens públicas oferecem soluções prontas, mas seus custos aumentam significativamente conforme a adoção cresce.

Além disso, treinar um modelo de IA também é um processo demorado, e as empresas interessadas em ter modelos personalizados de IA Generativa precisam lançar a solução em um curto espaço de tempo para melhorar seu desempenho e liberar recursos para tarefas mais importantes.

# Solução

Usando os servidores Edge da Lenovo com GPUs da NVIDIA, a Edgebricks oferece uma solução que permite às empresas construir, treinar e implantar modelos de IA/ML 10 vezes mais rápido no local de sua escolha. Isso proporciona às empresas a facilidade de uso de uma nuvem pública, mas com o controle e a flexibilidade de uma nuvem privada de IA/ML.

A solução de PaaS de IA/ML da Edgebricks é abrangente e de ponta a ponta, permitindo que as empresas simplesmente comecem executando seus pipelines. A plataforma concede acesso aos engenheiros de IA/ML para eles operarem de forma autônoma.

Oferecemos um modelo único em que você pode transferir estágios caros do pipeline que fazem o treinamento para infraestrutura no local ou em um colocation. Fornecemos um armazenamento de objetos integrado na nuvem da Edgebricks para armazenar esses dados e executar trabalhos de treinamento nos servidores com acesso rápido aos dados na mesma LAN. Não há necessidade de pagar altos custos de armazenamento ou acesso de dados na nuvem.

## Resultados

Usando a plataforma como serviço da Edgebricks, as empresas podem construir pipelines de IA/ML em questão de dias. A Edgebricks suporta os seguintes projetos de IA prontos para uso: detecção de objetos, processamento de linguagem natural, sistemas de recomendação e modelos de linguagem natural privados.

Uma vez que a infraestrutura está no lugar, a Edgebricks pode transformá-la em uma nuvem de IA/ML em um dia e ajudar a construir, treinar, testar e implantar modelos de IA sem a necessidade de muita expertise em DevOps. Os desenvolvedores têm um portal de autoatendimento para testar e implantar aplicativos. As equipes de DevOps recebem fluxos de trabalho de construção de pipeline automatizados, para que não precisem costurar tudo por conta própria.

Os três principais benefícios incluem:

- Construir, treinar e implantar modelos 10 vezes mais rápido em comparação com a execução por conta própria
- Obter aplicativos de ponta com automação para implantá-los
- Reduzir o custo total de propriedade em 80% em comparação com o uso de nuvens públicas

Além desses benefícios, os clientes também obtêm privacidade de dados e conformidade, pois podem implantar essa infraestrutura através de seu próprio firewall no local de sua escolha.

Benefícios adicionais incluem:

- Configurar infraestrutura e pipeline de IA/ML em algumas horas
- Reduzir o custo total de propriedade em 80%
- Oferecer autoatendimento à equipe de desenvolvimento de IA
- Fornecer controle, visibilidade e baixo OpEx à equipe de DevOps

## Arquitetura Validada

Esta solução utiliza os servidores ThinkEdge SE450 e SR650 V2 da Lenovo como a camada central da infraestrutura. Esses servidores terão CPU Intel Xeon com, em média, 16 e 32 núcleos por CPU, até 1TB de RAM, SSDs para armazenamento rápido e GPUs NVIDIA A100 Tensor Core para as workloads de IA/ML mais exigentes. Os clientes também receberão suporte e garantia de alta qualidade da Lenovo.

O software Edgebricks transformará esses servidores em uma nuvem de IA/ML. Os clientes podem, então, consumir essa nuvem usando um portal web simples. A Edgebricks fornece uma loja de aplicativos integrada com ferramentas de IA/ML e um construtor de pipelines para acelerar projetos de IA em uma empresa.



Há uma loja de aplicativos integrada com dezenas de ferramentas de IA/ML. A implantação e integração dessas ferramentas são automatizadas para construir vários fluxos de trabalho e pipelines conforme necessário. Isso inclui a automação de detecção de objetos, rotulagem de dados, processamento de linguagem natural, sistemas de recomendação e outros.

## Componentes de Design

Servidores	Outras Especificações	Acelerador	Caso de Uso	Software
ThinkEdge SE450	rede de vários gigabits e SSD local por host	NVIDIA A30	Workloads de IA/ML em Escala	Edgebricks PaaS de IA/ML
ThinkEdge SR650V2	rede de vários gigabits e SSD local por host	NVIDIA A100	Construção e Inferência de IA/ML	Edgebricks PaaS de IA/ML
ThinkEdge SE450	rede de vários gigabits e SSD local por host	-	Nuvem Privada Edge/Colo	Edgebricks Infraestrutura Edge/Nuvem

## Recursos

- Explore o Programa Lenovo AI Innovators
- Explore o Centro de Inovação e Briefing de HPC e IA da Lenovo
- Design Validado da Lenovo para Infraestrutura de IA em Servidores ThinkSystem
- Site da Edgebricks
- Edgebricks - Construa Pipelines de IA/ML em Minutos
- Aliança Lenovo-NVIDIA

## Por que escolher a Lenovo

Focada em uma visão audaciosa de fornecer tecnologia mais inteligente para todos, a Lenovo está desenvolvendo tecnologias que mudam o mundo, criando uma sociedade digital mais inclusiva, confiável e sustentável. Ao projetar, desenvolver e construir o portfólio mais completo de dispositivos e infraestrutura inteligentes do mundo, também lideramos uma Transformação Inteligente para criar experiências e oportunidades melhores para milhões de clientes ao redor do mundo.

## Por que escolher a NVIDIA

A NVIDIA foi pioneira na computação acelerada para enfrentar desafios que mais ninguém pode resolver. Nosso trabalho em IA e no metaverso está impactando profundamente a sociedade e transformando as maiores indústrias do mundo, desde jogos até robótica, carros autônomos até cuidados de saúde que salvam vidas, mudanças climáticas até mundos virtuais onde todos podemos nos conectar e criar.



© 2024 Lenovo. Todos os direitos reservados.

Disponibilidade: Ofertas, preços, especificações e disponibilidade podem ser alterados sem aviso prévio. A Lenovo não se responsabiliza por erros fotográficos ou tipográficos.

Garantia: Para obter uma cópia das garantias aplicáveis, escreva para Informações sobre Garantia da Lenovo, 1009 Think Place, Morrisville, NC, 27560. A Lenovo não faz representação ou garantia em relação a produtos ou serviços de terceiros.

Marcas Registradas: Lenovo, o logotipo da Lenovo, ThinkSystem, ThinkAgile são marcas comerciais ou marcas registradas da Lenovo. Microsoft e Windows são marcas registradas da Microsoft Corporation. Intel, o logotipo da Intel e Xeon são marcas comerciais da Intel Corporation ou de suas subsidiárias. NVIDIA, RTX, o logotipo da NVIDIA são marcas comerciais e/ou marcas registradas da NVIDIA Corporation nos EUA e em outros países.