Dell Al Factory

Solução de lA generativa Dell com AMD

Acelere a inovação, reduza os custos e proteja os dados com uma arquitetura modular e escalável para a complexa IA generativa.

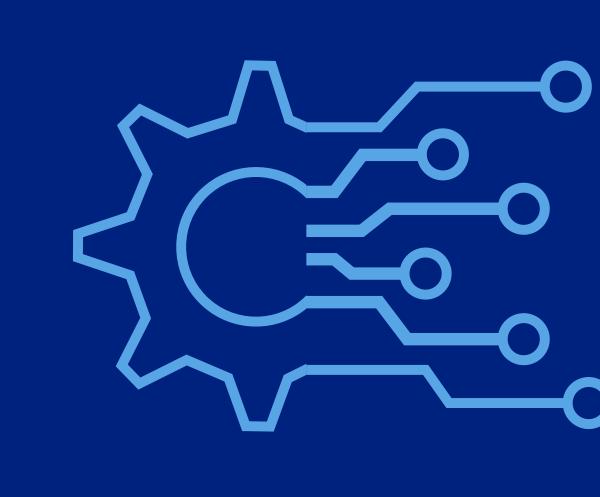




Os principais casos de uso exigem potência, flexibilidade e escala



Assistentes, chatbots e criação de conteúdo



Acelerador "as a service"



Geração aumentada de recuperação multimodal (RAG)



Simplificado

Simplifique as implementações da IA generativa com soluções validadas e comprovadas, respaldadas por mais de 340 mil horas de engenharia.

Otimize o desempenho

Acelerador de alto desempenho, arquitetura aberta e fabricações otimizadas para IA

A IA em todo lugar

Dados em qualquer lugar com flexibilidade de armazenamento em várias nuvens

Vários nós turnkey

Fundamentos comprovados de IA de pilha completa que proporcionam resultados mais rápidos



Hersonalizado

abertos impulsionam o desenvolvimento e as operações de IA.

Acelere o

O software de código aberto AMD ROCm™ e os ecossistemas

Inovar mais rápido Use software de código

aberto e ecossistemas para desenvolver aplicativos exclusivos.

desenvolvimento Aproveite as estruturas

padrão do setor com conjuntos de tecnologia flexíveis.

Ative seus dados

Execute com eficiência vários casos de uso de IA simultaneamente.



82% dos ITDMs preferem um modelo no local ou híbrido.3

Comece rapidamente Simplifique a **Provisionamento** automático Bases no local com conectividade

Seus dados determinam seus resultados. Proteja-a.

segurança de raiz de

confiança e controle total

AMD

ROCm

Fabrics seguros e

repletos de recursos com escalabilidade e fluxos de tráfego otimizados

Base de código aberto para implementação

e gerenciamento de clusters de alto desempenho

Estrutura de IA/ML baseada em padrões

Estrutura de IA do Acelerador AMD

Soluções de lA generativa da Dell com a AMD

Gerenciamento de infraestrutura Dell Omnia **PowerEdge** PowerEdge XE9680 com GPU AMD MI300x **PowerScale PowerSwitch Dell Professional Services**

Inferência

Personalizar

parâmetro 70B em um único acelerador AMD Instinct™

Execute um modelo de

MI300X.4

Aumentar Incorpore seus dados ao Implante e faça o ajuste fino de oito modelos 70B simultâneos em um único

processo generativo.

Obtenha diferenciação competitiva com uma solução aberta e comprovada que oferece aplicativos de IA seguros e no local em escala.

Dell PowerEdge XE9680.4

Saiba mais

¹Enterprise Strategy Group, <u>Maximizando o ROI da IA: As inferências no local</u> com a Dell Technologies podem ser 75% mais econômicas que as nuvens <u>públicas</u>, abril de 2024. ² Estimativa com base em uma análise da Dell de maio de 2024 que compara o tempo de configuração de um cluster Kubernetes de dois nós para um

instalação básica. O tempo real de configuração varia de acordo com a configuração da solução.

2023. ⁴Blog da Dell Technologies, <u>Silicon Diversity: Implementação da IA</u> generativa no PowerEdge XE9680 com aceleradores AMD Instinct MI300X,

³ Dell Technologies <u>Generative Al Pulse Survey</u>, de agosto e setembro de

LLM de uso geral usando scripts automatizados com a implementação manual de um projeto comum. O tempo de configuração inclui apenas a maio de 2024. Direitos autorais © 2024 Dell Inc. ou suas subsidiárias. Todos os direitos reservados. Dell e outras marcas comerciais pertencem à Dell Inc. ou às suas

subsidiárias. AMD, o logotipo da seta AMD, ROCm™, Instinct™ e suas combinações são marcas comerciais da Advanced Micro Devices, Inc. PyTorch® é uma

marca comercial ou marca registrada da PyTorch ou dos licenciadores da PyTorch. TensorFlow™ e quaisquer marcas relacionadas são marcas registradas da