Dell AI Factory

Solución de IA generativa de Dell con AMD

Acelere la innovación, reduzca los costes y proteja los datos con una arquitectura ampliable y modular para acoger una IA generativa compleja.

75 % más de rentabilidad en inferencia que la IaaS de cloud pública[1]

86 % menos de tiempo de rentabilización[2]

Los casos prácticos clave exigen potencia, flexibilidad y capacidad de ampliación

- Asistentes, chatbots y creación de contenido
- Acelerador como servicio
- Generación aumentada de recuperación (RAG) multimodal

Sencillez

Optimice las implementaciones de IA generativa con soluciones validadas, probadas y respaldadas por más de 340 000 horas de trabajo de ingeniería.

Optimice el rendimiento
Acelerador de alto rendimiento, arquitectura abierta y tejidos optimizados para IA

IA en cualquier lugar
En todos los datos gracias a la flexibilidad del almacenamiento multicloud

De múltiples nodos lista para usar
Bases de IA de pila completa probadas que ofrecen resultados más rápidos

Personalización

El software de código abierto AMD ROCm™ y los ecosistemas abiertos impulsan el desarrollo y las operaciones de IA.

Innove más rápido
Utilice software y ecosistemas de código abierto para desarrollar aplicaciones únicas.

Impulse el desarrollo
Aproveche las infraestructuras estándar del sector con pilas de tecnología flexibles.

Active los datos
Ponga en marcha varios casos prácticos de IA a la vez de forma eficiente.

Confianza

El 82 % de los responsables de la toma de decisiones de TI prefieren un modelo híbrido o en las propias instalaciones[3]. Los datos determinan los resultados. Protéjalos.

Comience rápido
Bases en las instalaciones con seguridad de raíz de confianza y control total

Optimice la conectividad
Tejidos seguros y repletos de funciones con capacidad de ampliación y flujos de tráfico optimizados

Automatice el aprovisionamiento
Base de código abierto para implementar y gestionar clústeres de alto rendimiento

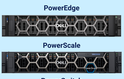
Soluciones de IA generativa de Dell con AMD

Dell Omnia

- Infraestructura de IA/ML basada en estándares
- AMD ROCm — Infraestructura de IA de AMD Accelerator
- Administración de la infraestructura

PowerEdge XE9680 con GPU AMD MI300x

PowerEdge



PowerScale

PowerSwitch

- Servicios profesionales Dell

Inferencia

Ejecute un modelo de 70 000 millones de parámetros en un solo acelerador AMD Instinct™ MI300X.[4]

Personalización

Implemente y ajuste con precisión ocho modelos simultáneos de 70 000 millones en un solo Dell PowerEdge XE9680[4].

Ampliación

Incorpore sus datos al proceso generativo.

Consiga diferenciarse de la competencia con una solución abierta y probada que ofrece aplicaciones de IA seguras en las instalaciones a escala.

Más información

[1] Enterprise Strategy Group, Maximización del ROI de la IA: la inferencia en las instalaciones con Dell Technologies puede ser hasta un 75 % más rentable que la cloud pública, abril de 2024.
[2] Estimación basada en un análisis de Dell realizado en mayo de 2024 en el que se compara el tiempo necesario para configurar un clúster de Kubernetes de 2 nodos para LLM de uso general mediante herramientas de script automatizados frente a la implementación manual de un diseño común. El tiempo de configuración incluye únicamente la instalación básica. El tiempo de configuración real variará en función de la configuración de la solución.
[3] Dell Technologies, Generative AI Pulse Survey, agosto y septiembre de 2023.
[4] Blog de Dell Technologies, Silicon Diversity: Deploy GenAI on the PowerEdge XE9680 with AMD Instinct MI300X Accelerators, mayo de 2024.

Copyright © 2024 Dell Inc. o sus filiales. Todos los derechos reservados. Dell y otras marcas comerciales pertenecen a Dell Inc. o sus filiales. AMD, el logotipo de AMD Arrow, ROCm™, Instinct™ y sus combinaciones son marcas comerciales de Advanced Micro Devices, Inc. PyTorch® es una marca comercial o una marca registrada de PyTorch o los licenciantes de PyTorch. TensorFlow™ y sus marcas relacionadas son marcas registradas de Google Inc. Las demás marcas comerciales pueden ser propiedad de sus respectivos propietarios. Infografía publicada en EE. UU. en octubre de 2024.

Dell Technologies considera que la información de este documento es precisa en el momento de su publicación. La información está sujeta a cambios sin previo aviso.