

Lenovo y NVIDIA impulsan la solución de Plataforma como Servicio (PaaS) de Modelos de IA de Edgebricks

La plataforma de Edgebricks proporciona una Infraestructura de Aprendizaje Automático Totalmente Automatizada para construir, entrenar e implementar modelos de IA 10 veces más rápido

Resumen

Hoy en día, el 89% de las empresas buscan hacer más con la adopción de la Inteligencia Artificial y el Aprendizaje Automático en sus productos existentes, procesos y trabajo diario. La IA generativa está causando un aumento adicional en las aplicaciones potenciales en 100 veces a medida que pasamos de la IA basada en la clasificación a la generación de contenido para imágenes, videos, código, texto y otros tipos de medios.

La construcción y ejecución de modelos GPT privados son esenciales para muchas industrias debido a cuestiones de cumplimiento, privacidad de datos y costos. El Aprendizaje Automático también es la workload más costosa de alojar en una nube pública, ya que requiere alta capacidad de cómputo, memoria, ancho de banda de red y almacenamiento.

La plataforma de IA/ML de Edgebricks como servicio ayuda a las empresas a construir, entrenar e implementar IA/ML 10 veces más rápido utilizando los servidores Edge de Lenovo con GPU de NVIDIA en el lugar que elijan. Así obtienen la facilidad de uso de una nube pública pero con el control y la flexibilidad de una nube privada de IA/ML.

Desafío

El crecimiento de la IA Empresarial y la adopción de la IA Generativa están llevando a las empresas a adoptar IA/ML para mantenerse competitivas, aumentar la productividad de sus equipos y reducir costos. Deben construir la infraestructura de IA/ML por sí mismas o aprovechar alguna solución de talla única en nubes públicas. Sin embargo, la mayoría menciona la contratación del talento adecuado para IA/ML como uno de sus mayores problemas.

El conjunto general de herramientas y el panorama están evolucionando rápidamente, y no hay una manera fácil de construir infraestructura y tuberías de IA/ML sin un alto costo y experiencia. Las nubes públicas ofrecen soluciones listas, pero sus costos aumentan significativamente a medida que aumenta la adopción.

Además, entrenar un modelo de IA también es un proceso que consume mucho tiempo y las empresas interesadas en tener modelos de IA Generativa personalizados se ven presionadas a lanzar la solución en un corto período de tiempo para mejorar su rendimiento y liberar recursos para tareas más importantes.

The Lenovo logo consists of the word "Lenovo" in white text on a red rectangular background.The NVIDIA logo features a green square with a white eye-like shape inside, followed by the word "NVIDIA" in black text.

Solución

Utilizando los servidores Edge de Lenovo con GPU de NVIDIA, Edgebricks ofrece una solución que permite a las empresas construir, entrenar e implementar modelos de IA/ML 10 veces más rápido en el lugar que elijan. Esto les brinda a las empresas la facilidad de uso de una nube pública pero con el control y la flexibilidad de una nube privada de IA/ML.

La solución PaaS de IA/ML de Edgebricks es integral y de extremo a extremo, lo que permite a las empresas comenzar simplemente ejecutando sus tuberías. La plataforma otorga acceso a sus ingenieros de IA/ML, permitiéndoles operar de manera autoservicio.

Ofrecemos un modelo único donde puedes externalizar etapas costosas del pipeline que realizan el entrenamiento a infraestructura en las instalaciones o en un colocation. Proporcionamos un almacenamiento de objetos integrado en la nube de Edgebricks para almacenar estos datos y ejecutar trabajos de entrenamiento en los servidores con acceso rápido a los datos en la misma LAN. No es necesario pagar altos costos de almacenamiento o acceso de datos en la nube.

Resultados

Utilizando la plataforma como servicio de Edgebricks, las empresas pueden construir tuberías de IA/ML en cuestión de días. Edgebricks admite los siguientes proyectos de IA listos para usar: detección de objetos, procesamiento del lenguaje natural, sistemas de recomendación y modelos de lenguaje natural privados.

Una vez que la infraestructura está en su lugar, Edgebricks puede convertirla en una nube de IA/ML en un día y ayudar a construir, entrenar, probar e implementar modelos de IA sin que las empresas necesiten mucha experiencia en DevOps. Los desarrolladores tienen un portal de autoservicio para probar e implementar aplicaciones. Los equipos de DevOps obtienen flujos de trabajo de construcción de tuberías automatizados, por lo que no tienen que unir todo por sí mismos.

Los tres principales beneficios incluyen:

- Construir, entrenar e implementar modelos 10 veces más rápido en comparación con hacerlo por sí mismos
- Obtener aplicaciones de primera con automatización para implementarlas
- Reducir el costo total de propiedad en un 80% en comparación con el uso de nubes públicas

Además de estos beneficios, los clientes también obtienen privacidad de datos y cumplimiento, ya que pueden implementar esta infraestructura detrás de su propio firewall en el lugar que elijan.

Otros beneficios incluyen:

- Configurar infraestructura y tubería de IA/ML en pocas horas
- Reducir el costo total de propiedad en un 80%
- Proporcionar autoservicio al equipo de desarrollo de IA
- Brindar control, visibilidad y bajos costos operativos al equipo de DevOps

Arquitectura Validada

Esta solución utiliza los servidores ThinkEdge SE450 y SR650 V2 de Lenovo como la capa central de la infraestructura. Estos servidores tendrán CPU Intel Xeon con entre 16 y 32 núcleos por CPU, hasta 1TB de RAM, SSD para almacenamiento rápido y GPU NVIDIA A100 Tensor Core para las workloads de IA/ML más exigentes. Los clientes también recibirán el mejor soporte y garantía de Lenovo.

El software Edgebricks convertirá estos servidores en una nube de IA/ML. Los clientes pueden consumir esta nube mediante un sencillo portal web. Edgebricks proporciona una tienda de aplicaciones integrada con herramientas de IA/ML y un constructor de tuberías para acelerar proyectos de IA en una empresa.



Hay una tienda de aplicaciones integrada con docenas de herramientas de IA/ML. La implementación e integración de estas herramientas están automatizadas, para construir varios flujos de trabajo y tuberías según sea necesario. Estas incluyen la automatización de la detección de objetos, el etiquetado de datos, el procesamiento del lenguaje natural, sistemas de recomendación y otros.

Componentes de Diseño

Servidores	Otras Especificaciones	Acelerador	Caso de Uso	Software
ThinkEdge SE450	red de varios gigabits y SSD local por host	NVIDIA A30	Workloads de IA/ML a Escala	Edgebricks PaaS de IA/ML
ThinkEdge SR650V2	red de varios gigabits y SSD local por host	NVIDIA A100	Construcción e Inferencia de IA/ML	Edgebricks PaaS de IA/ML
ThinkEdge SE450	red de varios gigabits y SSD local por host	-	Nube Privada Edge/Colo	Edgebricks Infraestructura Edge/Nube

Recursos

- Explore el Programa de Innovadores en IA de Lenovo
- Explore el Centro de Innovación y Presentación de HPC y IA de Lenovo
- Diseño Validado de Lenovo para Infraestructura de IA en Servidores ThinkSystem
- Sitio web de Edgebricks
- Edgebricks - Construya Pipelines de IA/ML en Minutos
- Alianza Lenovo-NVIDIA

¿Por qué elegir Lenovo?

Centrada en una visión audaz de ofrecer tecnología más inteligente para todos, Lenovo está desarrollando tecnologías que cambian el mundo, creando una sociedad digital más inclusiva, confiable y sostenible. Al diseñar, ingeniar y construir el portafolio más completo de dispositivos e infraestructura inteligente del mundo, también lideramos una Transformación Inteligente para crear mejores experiencias y oportunidades para millones de clientes en todo el mundo.

¿Por qué elegir NVIDIA?

NVIDIA fue pionera en la computación acelerada para abordar desafíos que nadie más puede resolver. Nuestro trabajo en IA y el metaverso está impactando profundamente en la sociedad y transformando las industrias más grandes del mundo, desde juegos hasta robótica, autos autónomos hasta atención médica que salva vidas, cambio climático hasta mundos virtuales donde todos podemos conectarnos y crear.



© 2024 Lenovo. Todos los derechos reservados.

Disponibilidad: Las ofertas, precios, especificaciones y disponibilidad pueden cambiar sin previo aviso. Lenovo no se hace responsable de errores fotográficos o tipográficos.

Garantía: Para obtener una copia de las garantías aplicables, escriba a Información de Garantía de Lenovo, 1009 Think Place, Morrisville, NC, 27560. Lenovo no hace representación ni garantía con respecto a productos o servicios de terceros.

Marcas comerciales: Lenovo, el logotipo de Lenovo, ThinkSystem, ThinkAgile son marcas comerciales o marcas registradas de Lenovo. Microsoft y Windows son marcas registradas de Microsoft Corporation. Intel, el logotipo de Intel y Xeon son marcas comerciales de Intel Corporation o de sus subsidiarias. NVIDIA, RTX, el logotipo de NVIDIA son marcas comerciales y/o marcas registradas de NVIDIA Corporation en los EE. UU. y otros países.