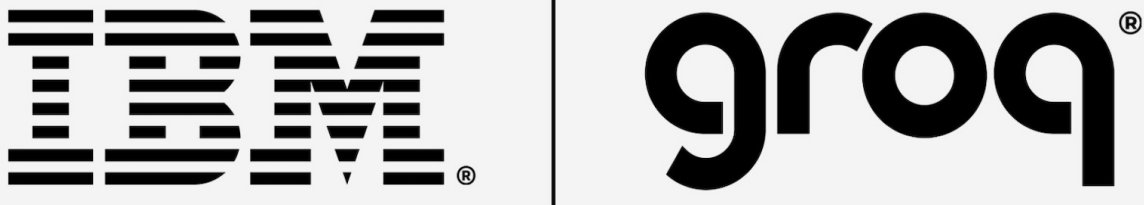


IBM y Groq se asocian para acelerar la implementación de IA empresarial con velocidad y escala

La alianza busca ofrecer capacidades de IA agente más rápidas a través de IBM watsonx Orchestrate y la tecnología de Groq, permitiendo que los clientes empresariales tomen acciones inmediatas sobre flujos de trabajo complejos



ARMONK, N.Y. y MOUNTAIN VIEW, Calif. – 20 de Octubre de 2025 – IBM (NYSE: IBM) y Groq anuncian hoy una alianza estratégica de comercialización y tecnología, diseñada para ofrecer a los clientes acceso inmediato a la tecnología de inferencia de Groq, GroqCloud, a través de watsonx Orchestrate. Esta colaboración proporciona capacidades de inferencia de IA de alta velocidad, con un costo que acelera la adopción de agentes de IA.

Como parte de la alianza, Groq e IBM planean integrar y mejorar la tecnología de código abierto vLLM de Red Hat con la arquitectura LPU de Groq. También está previsto que los modelos Granite de IBM sean compatibles con GroqCloud para los clientes de IBM.

Las empresas que están migrando de proyectos piloto a producción con agentes de IA aún enfrentan desafíos relacionados con la velocidad, el costo y la confiabilidad, especialmente en sectores críticos como salud, finanzas, gobierno, comercio minorista y manufactura. Esta alianza combina la velocidad de inferencia, la eficiencia de costos y el acceso a los modelos de código abierto más recientes de Groq con la orquestación de agentes de IA de IBM, ofreciendo la infraestructura necesaria para escalar a nivel empresarial.

Impulsado por su LPU personalizada, GroqCloud ofrece inferencia más de cinco veces más rápida y eficiente en términos de costo que los sistemas tradicionales basados en GPU. El resultado es una latencia consistentemente baja y un rendimiento confiable, incluso cuando las cargas de trabajo se escalan globalmente — algo especialmente poderoso para agentes de IA en industrias reguladas.

Por ejemplo, los clientes de IBM en el sector salud reciben miles de preguntas complejas de pacientes simultáneamente. Con Groq, los agentes de IA de IBM pueden analizar información en tiempo real y proporcionar respuestas precisas de inmediato, mejorando la experiencia del cliente y permitiendo que las organizaciones tomen decisiones más rápidas e inteligentes.

Esta tecnología también se está aplicando en industrias no reguladas. Clientes de IBM en los sectores de comercio minorista y bienes de consumo están utilizando Groq para agentes de recursos humanos, ayudando a automatizar procesos y aumentar la

productividad de los empleados.

“Muchas organizaciones empresariales tienen diversas opciones para la inferencia de IA durante la fase de experimentación, pero cuando quieren pasar a producción, necesitan garantizar que los flujos de trabajo complejos se implementen con éxito para asegurar experiencias de alta calidad”, dijo Rob Thomas, vicepresidente sénior de Software y director comercial de IBM. “Nuestra alianza con Groq refuerza el compromiso de IBM de ofrecer a los clientes las tecnologías más avanzadas para implementar IA y generar valor para los negocios.”

“Con la velocidad de Groq y la experiencia empresarial de IBM, estamos haciendo realidad la IA agente para los negocios. Juntos, estamos permitiendo que las organizaciones desbloqueen todo el potencial de las respuestas impulsadas por IA con el rendimiento necesario para escalar”, dijo Jonathan Ross, CEO y fundador de Groq. “Más allá de la velocidad y la resiliencia, esta alianza trata de transformar la forma en que las empresas trabajan con IA — pasando de la experimentación a la adopción empresarial con confianza, y abriendo camino a nuevos estándares en los que la IA puede actuar instantáneamente y aprender continuamente.”

IBM ofrecerá acceso inmediato a las capacidades de GroqCloud, y los equipos conjuntos se enfocarán en entregar las siguientes funcionalidades a los clientes de IBM, incluyendo:

- Inferencia de alta velocidad y alto rendimiento, que desbloquea todo el potencial de los modelos de IA y de la IA agente, impulsando casos de uso como atención al cliente, soporte a empleados y aumento de la productividad.
- Implementación de IA con enfoque en seguridad y privacidad, diseñada para cumplir con los requisitos regulatorios y de seguridad más estrictos, permitiendo la ejecución eficaz de flujos de trabajo complejos.
- Integración fluida con el producto de IA agente de IBM, watsonx Orchestrate, ofreciendo a los clientes flexibilidad para adoptar estándares de IA agente desarrollados específicamente para diferentes casos de uso.

La asociación también planea integrar y mejorar la tecnología vLLM de código abierto de Red Hat con la arquitectura LPU de Groq para ofrecer diferentes enfoques a los desafíos comunes de la IA que enfrentan los desarrolladores durante la inferencia. Se espera que la solución permita a watsonx aprovechar estas capacidades de manera familiar y que los clientes puedan seguir utilizando sus herramientas preferidas mientras aceleran la inferencia con GroqCloud. Esta integración abordará necesidades clave de los desarrolladores de IA, incluyendo la orquestación de inferencias, el balanceo de carga y la aceleración por hardware, optimizando en última instancia el proceso de inferencia.

Juntos, IBM y Groq ofrecen un acceso mejorado al potencial completo de la IA empresarial: rápida, inteligente y diseñada para generar impacto en el mundo real.

Las declaraciones sobre la dirección futura e intenciones de IBM y Groq están sujetas a cambios o retiro sin previo aviso, y representan únicamente metas y objetivos.

Acerca de IBM

IBM es un proveedor líder global en nube híbrida, inteligencia artificial y consultoría especializada. Ayudamos a clientes en más de 175 países a aprovechar el valor de sus datos, optimizar procesos empresariales, reducir costos y obtener ventajas competitivas en sus sectores. Miles de gobiernos y empresas en áreas críticas de infraestructura — como servicios financieros, telecomunicaciones y salud — confían en la plataforma de nube híbrida de IBM y en Red Hat OpenShift para llevar a cabo sus

transformaciones digitales de manera rápida, eficiente y segura. Las innovaciones revolucionarias de IBM en IA, computación cuántica, soluciones en la nube específicas por industria y servicios de consultoría ofrecen opciones abiertas y flexibles a nuestros clientes. Todo esto está respaldado por el compromiso histórico de IBM con la confianza, la transparencia, la responsabilidad, la inclusión y la excelencia en el servicio. Más información:www.ibm.com.
