

IBM z17: El primer Mainframe totalmente diseñado para la era de la IA

Las nuevas innovaciones desbloquean las capacidades de la IA a escala empresarial, incluidos los grandes modelos lingüísticos y la IA generativa.

Reimaginación de la experiencia del usuario mediante nuevos asistentes y agentes.

IBM (NYSE: [IBM](#)) ha anunciado hoy el [IBM z17](#), la próxima generación del icónico mainframe de la compañía, totalmente diseñado con capacidades de IA en hardware, software y operaciones de sistemas. Equipado con el nuevo procesador IBM Telum® II, IBM z17 amplía las capacidades del sistema más allá de la IA transaccional para permitir nuevas cargas de trabajo.

IBM Z está construido para redefinir la IA a escala, posicionando a las empresas para puntuar el 100% de sus transacciones en tiempo real.¹ z17 permite a las empresas impulsar la innovación y hacer más, incluyendo la capacidad de procesar un 50% más de operaciones de inferencia de IA por día que z16.² El nuevo IBM z17 está construido para aumentar el valor de negocio en todas las industrias con una amplia gama de más de 250 casos de uso de IA, tales como la mitigación del riesgo de préstamos, la gestión de servicios de chatbot, el apoyo al análisis de imágenes médicas o la obstaculización de la delincuencia minorista, entre otros.

IBM z17 es la culminación de cinco años de diseño y desarrollo que incluyeron la presentación de más de 300 solicitudes de patentes presentadas ante la Oficina de Patentes y Marcas de Estados Unidos. Diseñado con el aporte directo de más de 100 clientes y en estrecha colaboración con los equipos de Investigación y Software de IBM, el nuevo sistema introduce capacidades de IA multimodelo, nuevas funciones de seguridad para proteger los datos y herramientas que aprovechan la IA para mejorar la usabilidad y la gestión del sistema:

- **Llevar la IA a los datos:** las capacidades de inferencia de la IA de z17 están impulsadas por un acelerador de IA en chip de segunda generación integrado en el procesador IBM Telum® II, que ofrece una mayor frecuencia, capacidad de cálculo y un aumento del 40% de la memoria «caché», lo que permite realizar más de 450.000 millones de operaciones de inferencia en un día y un tiempo de respuesta de un milisegundo.²
- **Ampliar la aceleración para la IA:** el acelerador IBM Spyre™ Accelerator, que se espera que esté disponible en el cuarto trimestre de 2025 a través de una tarjeta PCIe, proporcionará capacidades de computación de IA adicionales para complementar el procesador Telum® II. Juntos, crearán entornos optimizados para dar soporte a métodos multimodelo de IA. El Spyre™ Accelerator está especialmente diseñado para llevar las capacidades de IA generativa al mainframe, incluidos los asistentes de ejecución, aprovechando los datos empresariales contenidos en el sistema.
- **Aprovechar la IA para mejorar la experiencia del usuario:** z17 está diseñado para reforzar las habilidades y la eficiencia de los desarrolladores y las operaciones de TI con el uso de asistentes de IA y agentes de IA, incluyendo [IBM watsonx Code Assistant para Z](#) e [IBM watsonx Assistant para Z](#). Además, por primera vez, watsonx Assistant para Z se integrará con Z Operations Unite para proporcionar detección y resolución de incidentes basada en chat de IA utilizando datos de sistemas en vivo.

“La industria está aprendiendo rápidamente que la IA solo será tan valiosa como la infraestructura en la que se ejecute», dijo Ross Mauri, director general de IBM Z y LinuxONE, IBM. “Con z17, estamos llevando la IA al núcleo de la empresa con el software, la potencia de procesamiento y el almacenamiento para que la IA sea operativa rápidamente. Además, las

organizaciones pueden poner sus vastos y no utilizados almacenes de datos empresariales a trabajar con IA de forma segura y rentable."

Integración total de Hardware y Software

IBM z17 es un sistema diseñado desde cero para integrarse plenamente en entornos híbridos mediante una estrecha unión de innovaciones de hardware, capacidades de software para IA y un amplio soporte para herramientas y estándares abiertos. Esto permite un rendimiento y una fiabilidad diferenciados a la vez que reimagina cómo los desarrolladores y los operadores de sistemas interactúan con IBM Z y lo gestionan:

- **Sistema operativo para IA** - IBM también presentó un avance de z/OS 3.2, la próxima versión de su sistema operativo estrella para IBM Z, cuyo lanzamiento está previsto para el tercer trimestre de 2025. z/OS 3.2 está diseñado para soportar capacidades de IA acelerada por hardware en todo el sistema y conocimientos operativos de IA para las capacidades de gestión del sistema. Además, z/OS 3.2 ofrecerá soporte para métodos modernos de acceso a datos, bases de datos NoSQL y procesamiento de datos en la nube híbrida. Estas nuevas capacidades ayudarán al software de IA a acceder a un conjunto más amplio de datos empresariales y a obtener información empresarial predictiva.
- **Operaciones unificadas de TI** - También se ha anunciado hoy IBM Z Operations Unite, que reúne métricas de rendimiento clave y registros de múltiples fuentes a través de IBM Z, en formato OpenTelemetry, para agilizar las operaciones de IBM Z con IA. La nueva solución está diseñada para acelerar el tiempo de detección de anomalías, aislar el impacto de posibles incidentes y reducir el tiempo de resolución. Utilizada junto con IBM Concert, los equipos de operaciones pueden beneficiarse de la correlación inteligente de datos operativos en toda la empresa. IBM Operations Unite estará disponible de forma general en mayo de 2025.
- **Acelerador de IA para la eficiencia empresarial** - Con las opciones de ampliación del acelerador IBM Spyre™ Accelerator, que se espera que estén disponibles en el cuarto trimestre de 2025 a través de una tarjeta PCIe, IBM z17 busca transformar la experiencia del usuario en la plataforma. Los clientes podrán ejecutar el creciente catálogo de asistentes y agentes de IBM, basados en los modelos Granite de IBM, de forma nativa en z17 sin asumir el riesgo añadido asociado a mover datos o lógica de negocio sensible fuera de la plataforma. Juntas, estas soluciones están diseñadas para formar una estructura optimizada que permita a los clientes aumentar la productividad con seguridad y escalabilidad.

Construido para la resiliencia: Seguridad y Ciberseguridad en el centro

IBM z17 amplía el historial de sólidas capacidades de seguridad y resistencia de la plataforma. Los nuevos desarrollos en IA han permitido el despliegue de inteligencia añadida en esta área cada vez más importante para los clientes, ya que cada día aparecen nuevas amenazas. Esto incluye varias capacidades nuevas, entre ellas:

- **Protección de credenciales:** las funciones de HashiCorp, una empresa de IBM, anunciadas en marzo, ya están disponibles en IBM Z para ayudar a estandarizar la gestión de credenciales en la nube híbrida. IBM Vault utiliza la seguridad basada en identidades para autenticar y autorizar el acceso a credenciales, certificados, claves, tokens y otros datos sensibles. Con la incorporación de IBM Vault, los clientes pueden disponer de una única solución para ayudar a proteger las cargas de trabajo críticas gestionando todo el ciclo de vida de las credenciales en todo su stack de tecnología.
- **Seguridad de datos impulsada por IA:** IBM propone ofrecer nuevas capacidades para descubrir y clasificar datos sensibles en la plataforma. Esto aprovecharía Telum® II y utilizaría el procesamiento del lenguaje natural para poder

identificar y proteger los datos de misión crítica. Además, nuestra última solución de seguridad basada en IA, [IBM Threat Detection for z/OS](#), está diseñada para detectar e identificar anomalías potencialmente maliciosas que podrían ser el resultado de un ciberataque.

IBM amplía el soporte de IA a IBM z17

La experiencia de soporte integral y a medida de IBM ayuda a los clientes de IBM Z a satisfacer demandas que van más allá del mantenimiento tradicional. Proporcionado por IBM Technology Lifecycle Services, IBM Support for z17 ayuda a los clientes a optimizar sus entornos para obtener el máximo rendimiento y hacer frente a los riesgos y las interrupciones de las operaciones de misión crítica. Los procesos de IA de IBM agilizan la resolución de incidencias y ayudan a mejorar el tiempo de resolución de casos, basados en IBM watsonx, que ahora dan soporte a los sistemas IBM Z.

IBM ofrece almacenamiento seguro y ágil

[IBM Storage DS8000](#) desempeña un papel clave como solución de almacenamiento integrada para IBM Z. La última generación de IBM Storage DS8000 ([10th Generation](#)) está diseñada para aprovechar toda la potencia de IBM z17, proporcionando a las organizaciones acceso a cargas de trabajo críticas, un rendimiento de datos coherente y optimizado, y una arquitectura modular para adoptar las últimas tecnologías respaldadas por la investigación de IBM para impulsar el crecimiento empresarial al tiempo que se monetizan los datos. Juntos, IBM Z e IBM Storage ofrecen una infraestructura moderna que proporciona una plataforma segura y ágil para cargas de trabajo de misión crítica.

Disponibilidad

IBM z17 estará disponible el 18 de junio de 2025. Para obtener más información, visite ibm.com/mx-es/z17. Se espera que IBM Spyre™ Accelerator esté disponible a partir del cuarto trimestre de 2025.

Las declaraciones relativas a la dirección e intenciones futuras de IBM están sujetas a cambios o retirada sin previo aviso, y representan únicamente metas y objetivos.

Acerca de IBM

IBM es un proveedor líder de [nube híbrida](#) global e [inteligencia artificial](#), y experiencia en [consultoría](#). Ayudamos a clientes en más de 175 países a capitalizar el conocimiento de sus datos, optimizar los procesos de negocios, reducir costos y obtener una ventaja competitiva en sus industrias. Miles de entidades gubernamentales y corporativas en áreas de infraestructura crítica como servicios financieros, telecomunicaciones y atención médica confían en la plataforma de nube híbrida de IBM y Red Hat OpenShift para lograr sus transformaciones digitales de manera rápida, eficiente y segura. Las innovaciones revolucionarias de IBM en inteligencia artificial, computación cuántica, soluciones de nube específicas de la industria y consultoría ofrecen opciones abiertas y flexibles a nuestros clientes. Todo esto está respaldado por el legendario compromiso de IBM con la confianza, la transparencia, la responsabilidad, la inclusión y el servicio. Visite www.ibm.com/mx-es para obtener más información.

Recursos Adicionales

- Software [blog](#)
- z17 Support [blog](#)
- Research [blog](#)

Media Contact:

Chase Skinner

IBM Communications

chase.skinner@ibm.com

Aishwerya Paul

IBM Communications

aish.paul@ibm.com

1. Claim – Since the introduction of IBM z16 in 2022, IBM Z mainframes have been able to support AI inferencing directly in the mainframe, making it possible to score 100% of real-time transactions even in high-volume production environments.

Source - Celent report: '[Mitigating Fraud in The AI Age](#)' by Neil Katkov, 04/08/2025, commissioned by IBM

2. Claim - The percentage difference between IBM z17, that processes up to 450 billion inference operations per day with 1 ms response time using a Credit Card Fraud Detection Deep Learning model, and IBM z16, process up to 300 billion inference requests per day with 1ms response time using a Credit Card Fraud Detection model. For IBM z17 up to 450 billion inferences operations per day.

Disclaimer - For z17 performance result is extrapolated from IBM® internal tests running on IBM Systems Hardware of machine type 9175. The benchmark was executed with 1 thread performing local inference operations using a LSTM based synthetic Credit Card Fraud Detection model (<https://github.com/IBM/ai-on-z-fraud-detection>) to exploit the integrated Accelerator for AI. A batch size of 160 was used. IBM Systems Hardware configuration: 1 LPAR running Red Hat® Enterprise Linux® 9.4 with 6 IFLs (SMT), 128 GB memory. 1 LPAR with 2 CPs, 4 zIIPs and 256 GB memory running IBM z/OS® 3.1 with IBM z/OS Container Extensions (zCX) feature. Results may vary. For IBM z16, performance result is extrapolated from IBM internal tests running local inference operations in an IBM z16 LPAR with 48 IFLs and 128 GB memory on Ubuntu 20.04 (SMT mode) using a synthetic credit card fraud detection model (<https://github.com/IBM/ai-on-z-fraud-detection>) exploiting the IBM Integrated Accelerator for AI. The benchmark was running with 8 parallel threads each pinned to the first core of a different chip. The lscpu command was used to identify the core-chip topology. A batch size of 128 inference operations was used. Results were also reproduced using a z/OS V2R4 LPAR with 24 CPs and 256GB memory on IBM z16. The same credit card fraud detection model was used. The benchmark was executed with a single thread performing inference operations. A batch size of 128 inference operations was used. Results may vary.
