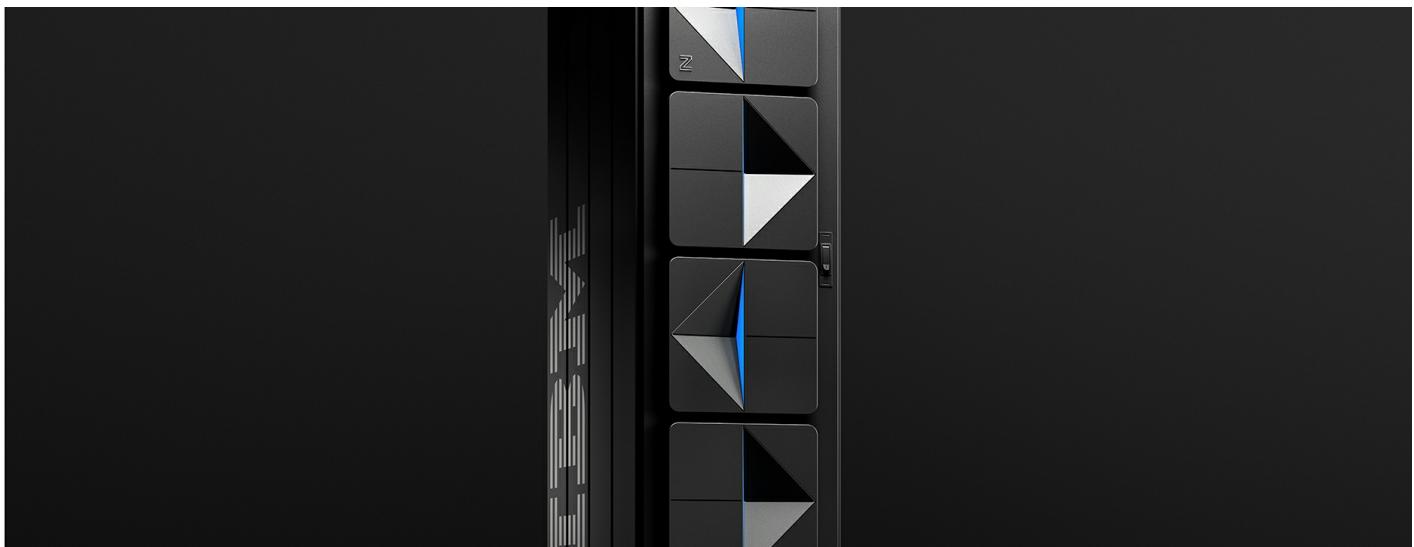


IBM z17: Der erste Mainframe, der vollständig für das KI-Zeitalter entwickelt wurde

Neue Innovationen eröffnen Möglichkeiten für KI im Unternehmensmaßstab, einschließlich großer Sprachmodelle (LLMs) und generative KI

Völlig neu konzipierte Nutzererfahrung durch neue KI-Assistenten und -Agenten



ARMONK, NY, 08. April 2025– IBM (NYSE: [IBM](#)) hat heute [IBM z17](#) angekündigt, die nächste Generation des legendären Mainframes des Unternehmens, der vollständig mit KI-Funktionen für Hardware-, Software- und das Betriebssystem ausgestattet ist. Angetrieben vom neuen IBM Telum® II Prozessor, erweitert IBM z17 die Funktionalität des Systems über transaktionale KI-Funktionen hinaus, um neue Workloads zu ermöglichen.

IBM Z wurde entwickelt, um KI in großem Maßstab neu zu definieren und Unternehmen in die Lage zu versetzen, 100 % ihrer Transaktionen in Echtzeit zu bewerten¹. z17 ermöglicht es Unternehmen, Innovationen voranzutreiben und mehr zu tun, einschließlich der Möglichkeit, 50 Prozent mehr KI-Inferenzoperationen pro Tag zu verarbeiten als z16². Die neue IBM z17 wurde entwickelt, um den geschäftlichen Mehrwert branchenübergreifend mit einer breiten Palette von mehr als 250 KI-Anwendungsfällen zu steigern, wie z. B. die Minderung von Kreditrisiken, die Verwaltung von Chatbot-Diensten, die Unterstützung medizinischer Bildanalysen oder die Eindämmung von Einzelhandelskriminalität.

IBM z17 ist das Ergebnis von fünf Jahren Design- und Entwicklungsarbeit, in denen mehr als 300 Patentanmeldungen beim US-Patent- und Markenamt eingereicht wurden. Entwickelt mit dem direkten Input von mehr als 100 Kund_innen und in enger Zusammenarbeit mit IBM Forschungs- und Softwareteams; Das neue System führt Multi-Modell KI-Funktionen und neue Sicherheitsfunktionen zum Schutz von Daten und Tools ein, die KI nutzen, um die Benutzerfreundlichkeit und Verwaltung des Systems zu verbessern:

- **KI in die Daten bringen**– Die KI-Inferenzfunktionen von z17 werden von einem On-Chip-KI-Accelerator der zweiten Generation unterstützt, der in den IBM Telum® II-Prozessor integriert ist und eine höhere Frequenz, Rechenkapazität und ein um 40 Prozent größeres Cache-Wachstum bietet. Das ermöglicht mehr als 450 Milliarden Inferenzoperationen an einem Tag und eine Reaktionszeit von einer Millisekunde.
- **Erweiterung der Beschleunigung für KI**– Der IBM Spyre™ Accelerator, der voraussichtlich im 4. Quartal 2025 als PCIe-

Karte verfügbar sein wird, wird zusätzliche KI-Rechenfunktionen bieten, um den Telum® II-Prozessor zu ergänzen. Gemeinsam werden sie optimierte Umgebungen schaffen, um die Multi-Modell-Methoden der KI zu unterstützen. Der Spyre™ Accelerator wurde speziell entwickelt, um generative KI-Funktionen auf den Mainframe zu bringen, einschließlich laufender Assistenten, die im System enthaltenen, unternehmensweiten Daten nutzen.

- **Nutzung von KI zur Verbesserung der Nutzererfahrung** – z17 wurde entwickelt, um die Fähigkeiten und die Effizienz von Entwickler_innen und IT-Abläufen durch den Einsatz von KI-Assistenten und KI-Agenten, einschließlich IBM watsonx Code Assistant für Z und IBM watsonx Assistant für Z, zu stärken. Darüber hinaus wird zum ersten Mal der watsonx Assistant for Z in Z Operations Unite integriert, um eine KI-Chat basierte Vorfallerkennung und -lösung anhand von Live-Systemdaten zu ermöglichen.

"Die Branche lernt schnell, dass KI nur so wertvoll ist wie die Infrastruktur, auf der sie läuft", sagte Ross Mauri, General Manager, IBM Z und LinuxONE, IBM. "Mit z17 bringen wir KI mit der Software, der Rechenleistung und dem Speicher in den Kern des Unternehmens, um KI schnell einsatzbereit zu machen. Darüber hinaus können Unternehmen ihre riesigen, ungenutzten Speicher an Unternehmensdaten auf sichere und kostengünstige Weise mit KI zusammenarbeiten lassen."

Vollständig in Hardware und Software integriert

IBM z17 ist ein System, das von Grund auf so konzipiert wurde, dass es sich vollständig in hybride Umgebungen integriert, indem es Hardware-Innovationen, Softwarefunktionen für KI und umfassende Unterstützung für offene Standards und Tools eng miteinander verbindet. Dies ermöglicht eine differenzierte Performance und Zuverlässigkeit und gestaltet gleichzeitig die Art und Weise neu, wie Entwickler_innen und Systembetreiber_innen mit IBM Z interagieren und diese verwalten, einschließlich:

- **Betriebssystem für KI** – IBM hat auch eine Preview auf z/OS 3.2 gegeben, die nächste Version seines Flaggschiff-Betriebssystems für IBM Z, die im dritten Quartal 2025 veröffentlicht werden soll. z/OS 3.2 wurde entwickelt, um hardwarebeschleunigte KI-Funktionen im gesamten System und betriebliche KI-Erkenntnisse für Systemverwaltungsfunktionen zu unterstützen. Darüber hinaus bietet z/OS 3.2 Unterstützung für moderne Datenzugriffsmethoden, NoSQL-Datenbanken und Hybrid-Cloud-Datenverarbeitung. Diese neuen Funktionen werden KI-Software dabei helfen, eine breitere Palette von Unternehmensdaten zu erschließen und vorausschauende Geschäftseinblicke abzuleiten.
- **Einheitlicher IT-Betrieb** – Ebenfalls angekündigt wurde heute IBM Z Operations Unite, welches wichtige Leistungskennzahlen und Protokolle aus mehreren Quellen in IBM Z im OpenTelemetry-Format zusammenführt, um den Betrieb von IBM Z mit KI zu optimieren. Die neue Lösung wurde entwickelt, um die Zeit bis zur Erkennung von Anomalien zu verkürzen, die Auswirkungen potenzieller Vorfälle zu verringern und die Lösungszeit zu verkürzen. In Verbindung mit IBM Concert können Operationsteams von der intelligenten Korrelation von Betriebsdaten im gesamten Unternehmen profitieren. IBM Operations Unite wird im Mai 2025 allgemein verfügbar sein.
- **KI-Accelerator für mehr Geschäfts-Effizienz** – Mit den Erweiterungsoptionen für den IBM Spyre™ Accelerator, der voraussichtlich im 4. Quartal 2025 als PCIe-Karte verfügbar sein wird, zielt IBM z17 darauf ab, die Nutzererfahrung auf der Plattform zu transformieren. Kund_innen werden in der Lage sein, den wachsenden Katalog von IBM Assistenten und Agenten, der auf IBMs Granite-Modellen basiert, nativ auf z17 auszuführen, ohne das zusätzliche Risiko einzugehen, welches mit der Verlagerung von Daten oder sensibler Geschäftslogik von der Plattform verbunden ist. Zusammen sind diese Lösungen so konzipiert, dass sie einen optimierten Stack bilden, der es den Kund_innen ermöglicht, mehr Produktivität mit Sicherheit und Skalierbarkeit zu erzielen.

Auf Ausfallsicherheit ausgelegt: Sicherheit und Cyberabwehr im Mittelpunkt

IBM z17 setzt die Geschichte der Plattform mit starken Sicherheits- und Resilienzfunktionen fort. Neue Entwicklungen in der KI haben die Nutzung zusätzlicher Informationen in diesem ständig wachsenden Bereich ermöglicht, der für Kund_innen von Bedeutung ist, da jeden Tag neue Bedrohungen auftreten. Dazu gehören mehrere neue Funktionen, einschließlich:

- **Verwaltung von Geschäftsgeheimnissen** – Die Funktionen von HashiCorp, einem IBM Unternehmen, die im März angekündigt wurden, sind jetzt auf IBM Z verfügbar, um die Standardisierung des Secrets-Managements in der Hybrid-Cloud zu unterstützen. IBM Vault verwendet identitätsbasierte Sicherheitsansätze, um den Zugriff auf Geschäftsgeheimnisse, Zertifikate, Keys, Token und andere vertrauliche Daten zu authentifizieren und zu autorisieren. Mit der Ergänzung von IBM Vault erhalten Kund_innen eine einheitliche Lösung, um kritische Workloads zu schützen. Indem sie über den gesamten Lebenszyklus hinweg ihre Geschäftsgeheimnisse über ihren gesamten IT-Bestand hinweg verwalten.
- **KI-gestützte Datensicherheit** – IBM beabsichtigt, neue Funktionen für die Erkennung und Klassifizierung sensibler Daten auf der Plattform bereitzustellen. Diese würden Telum® II einsetzen und die Verarbeitung natürlicher Sprache nutzen, sodass unternehmenskritische Daten identifiziert und geschützt werden können. Darüber hinaus wurde unsere neueste KI-gesteuerte Sicherheitslösung, IBM Threat Detection for z/OS, entwickelt, um potenziell bösartige Anomalien, die das Ergebnis eines Cyberangriffs sein könnten, zu erkennen und zu identifizieren.

IBM erweitert KI-gestützten Support auf IBM z17 aus

Der maßgeschneiderte, umfassende Support von IBM hilft IBM Z-Kund_innen, Anforderungen zu erfüllen, die über die herkömmliche Wartung hinausgehen. IBM Support for z17 wird von IBM Technology Lifecycle Services bereitgestellt und unterstützt Kund_innen bei der Optimierung ihrer Umgebungen für Spitzenleistung, um Risiken und Störungen bei unternehmenskritischen Vorgängen zu begegnen. Die KI-Prozesse von IBM, basierend auf IBM WatsonX, optimieren die Behebung von Vorfällen und tragen zur Verbesserung der Falllösungszeit bei und unterstützen jetzt IBM Z Systeme.

IBM bietet sicheren und flexiblen Speicher

IBM Storage DS8000 spielt als integrierte Speicherlösung für IBM Z eine Schlüsselrolle. Die neueste Generation von IBM Storage DS8000 ([10te Generation](#)) wurde entwickelt, um die volle Leistung von IBM z17 zu nutzen und Unternehmen Zugang zu kritischen Workloads, konsistenter und optimierter Datenleistung sowie einer modularen Architektur zu bieten, um die neuesten forschungsgestützten Technologien von IBM zu übernehmen, das Geschäftswachstum zu fördern und gleichzeitig Daten zu monetarisieren. Zusammen bieten IBM Z und IBM Storage eine moderne Infrastruktur, die eine sichere und agile Plattform für geschäftskritische Workloads bietet.

Verfügbarkeit

IBM z17 wird am 18. Juni 2025 allgemein verfügbar sein. Weitere Informationen finden Sie unter [IBM.com/z17](#). Der IBM Spyre™ Accelerator wird voraussichtlich ab dem 4. Quartal 2025 verfügbar sein.

Aussagen über die zukünftige Ausrichtung und Absicht von IBM können ohne vorherige Ankündigung geändert oder zurückgezogen werden und stellen nur Ziele dar.

Über IBM

IBM ist ein führender Anbieter von weltweiten Hybrid Cloud- und KI-Lösungen, sowie Expertenberatung. Wir helfen Kund_innen

in mehr als 175 Ländern, Erkenntnisse aus ihren Daten zu nutzen, Geschäftsprozesse zu optimieren, Kosten zu senken und sich einen Wettbewerbsvorteil in ihren Branchen zu verschaffen. Tausende von staatlichen und privaten Unternehmen in kritischen Infrastrukturbereichen wie Finanzdienstleistungen, Telekommunikation und Gesundheitswesen setzen auf IBMs Hybrid-Cloud-Plattform und Red Hat OpenShift, um ihre digitalen Transformationen schnell, effizient und sicher umzusetzen. IBMs bahnbrechende Innovationen in KI, Quantencomputing, branchenspezifischen Cloud-Lösungen und Beratung bieten unseren Kundinnen offene und flexible Optionen. All dies vor dem Hintergrund von IBMs langjährigem Engagement für Vertrauen, Transparenz, Verantwortung, Inklusivität und Service.

Weitere Quellen

- Ecosystem & Skills [blog](#)
- Software [blog](#)
- z17 Support [blog](#)
- Research [blog](#)

Medienkontakt:

Chase Skinner

IBM Communications

chase.skinner@ibm.com

Aishwarya Paul

IBM Communications

aish.paul@ibm.com

Petra Bernhardt

IBM DACH Communications

petra.bernhardt@de.ibm.com

1 Aussage – Seit der Einführung von IBM z16 im Jahr 2022 sind IBM Z-Mainframes in der Lage, KI-Inferenzen direkt im Mainframe zu unterstützen, wodurch es möglich ist, selbst in Produktionsumgebungen mit hohem Volumen 100 % der Echtzeittransaktionen zu erzielen.

Quelle - Celent berichtet: "[Eindämmung von Betrug im KI-Zeitalter](#)" von Neil Katkov, 04.08.2025, im Auftrag von IBM

2 Urheberrechtliche Aussage - Der prozentuale Unterschied zwischen IBM z17, die mithilfe eines Deep-Learning-Modells zur Kreditkartenbetrugserkennung bis zu 450 Milliarden Inferenzoperationen pro Tag mit einer Reaktionszeit von 1 ms verarbeitet, und IBM z16, die mithilfe eines Modells zur Kreditkartenbetrugserkennung bis zu 300 Milliarden Inferenzanfragen pro Tag mit einer Reaktionszeit von 1 ms verarbeitet. Bei IBM z17 sind es bis zu 450 Milliarden Inferenzoperationen pro Tag.

Haftungsausschluss - Das Leistungsergebnis für z17 wurde aus internen IBM®-Tests auf IBM Systems Hardware des Maschinentyps 9175 extrapoliert. Der Benchmark wurde mit einem Thread ausgeführt, der lokale Inferenzoperationen mithilfe eines LSTM-basierten synthetischen Kreditkartenbetrugserkennungsmodells (<https://github.com/IBM/ai-on-z-fraud-detection>) durchführte, um den integrierten Accelerator für KI zu nutzen. Es wurde eine Batchgröße von 160 verwendet. IBM Systems Hardware-Konfiguration: 1 LPAR mit Red Hat® Enterprise Linux® 9.4 mit 6 IFLs (SMT), 128 GB Speicher. 1 LPAR mit 2 CPs, 4 zIIPs und 256 GB Speicher mit IBM z/OS® 3.1 mit IBM z/OS Container Extensions (zCX). Ergebnisse können variieren. Für IBM z16 wird das Leistungsergebnis aus internen IBM-Tests extrapoliert, bei denen lokale Inferenzoperationen in einem IBM z16 LPAR mit 48 IFLs und 128 GB Speicher auf Ubuntu 20.04 (SMT-Modus) unter Verwendung eines synthetischen Modells zur Erkennung von Kreditkartenbetrug (<https://github.com/IBM/ai-on-z-fraud-detection>) ausgeführt wurden, das den IBM Integrated Accelerator for AI nutzt. Der Benchmark lief mit 8 parallelen Threads, die jeweils an den ersten Kern eines anderen Chips gebunden waren. Der Befehl lscpu wurde verwendet, um die Core-Chip-Topologie zu identifizieren. Es wurde eine Batchgröße von 128 Inferenzoperationen verwendet. Die Ergebnisse wurden auch unter Verwendung eines z/OS V2R4 LPAR mit 24 CPs und 256 GB Speicher auf IBM z16 reproduziert. Es wurde dasselbe Modell zur Erkennung von Kreditkartenbetrug verwendet. Der Benchmark wurde mit einem einzelnen Thread ausgeführt, der Inferenzoperationen durchführte. Es wurde eine Batchgröße von 128 Inferenzoperationen verwendet. Die Ergebnisse können variieren.

QUELLE IBM

- [New-IBM-z17-Mainframe](#)
- [IBM-Spyre-Accelerator-on-PCIe-Card](#)
- [IBM-Telum-II-Chip-IBM-Spyre-Accelerator-Chip](#)
- [IBM z17 Mainframe Video](#)