

[Comunicati stampa](#)

IBM z17, il primo mainframe completamente progettato per l'era dell'intelligenza artificiale

- **Innovazioni**, tra cui i Large Language Model e l'AI Generativa, consentono di sbloccare l'utilizzo dell'intelligenza artificiale su scala enterprise
- **Nuovi assistenti e agenti** per reinventare le esperienze utente



ARMONK, NY, 08 aprile 2025 - IBM (NYSE: IBM) annuncia oggi IBM z17, l'ultima generazione dell'iconico mainframe, completamente progettato integrando funzionalità di intelligenza artificiale nell'hardware, nel software e nel sistema operativo. Basato sul nuovo processore IBM Telum® II, IBM z17 espande le funzionalità del sistema oltre le funzionalità di intelligenza artificiale transazionale, abilitando la piattaforma alla gestione di nuovi carichi di lavoro.

IBM Z ridefinisce l'intelligenza artificiale su larga scala, permettendo alle aziende di valutare il 100% delle loro transazioni in tempo reale. Con z17 le aziende saranno in grado di guidare l'innovazione e fare molto di più, inclusa la capacità di elaborare il 50% in più di operazioni di inferenza AI al giorno rispetto a z16¹. Il nuovo IBM z17 può sostenere il business in tutti i settori di industria con oltre 250 casi d'uso sull'AI, come la mitigazione del rischio finanziario, la gestione dei servizi di chatbot, il supporto all'analisi delle immagini mediche o la prevenzione della micro-criminalità.

IBM z17 è il risultato di 5 anni di progettazione basata su oltre 300 brevetti registrati presso l'US Patent and Trademark Office. Con il contributo diretto di oltre 100 clienti che hanno collaborato con i team IBM Research e Software, il nuovo sistema introduce funzionalità AI multi-modello, nuove funzionalità di sicurezza per proteggere i dati e strumenti che utilizzano l'AI per migliorare l'usabilità e la gestione del sistema:

- **Portare l'AI nei dati** - le funzionalità di inferenza AI di z17 sono alimentate da un acceleratore AI on-chip di seconda

generazione integrato nel processore IBM Telum® II, caratterizzato da una maggiore frequenza, capacità di elaborazione, una crescita del 40 per cento nella cache, che consente oltre 450 miliardi di operazioni di inferenza al giorno e un tempo di risposta di un millisecondo.²

- **Aumentare l'accelerazione dell'AI** - IBM Spyre™ Accelerator, la cui disponibilità è prevista per il quarto trimestre 2025 tramite scheda PCIe, fornisce funzionalità di elaborazione AI aggiuntive che si integrano con il processore Telum® II. Insieme potranno creare ambienti ottimizzati per supportare implementazioni AI multi-modello. Spyre™ Accelerator è stato progettato appositamente per apportare funzionalità di AI generativa al mainframe, inclusa l'esecuzione di assistenti, utilizzando i dati aziendali contenuti nel sistema.
- **Migliorare l'esperienza utente con l'AI** - z17 è progettato per rafforzare le competenze e l'efficienza nello sviluppo e nelle operazioni IT con l'utilizzo di assistenti AI e agenti AI, tra cui IBM watsonx Code Assistant for Z e IBM watsonx Assistant for Z. Inoltre, per la prima volta watsonx Assistant for Z sarà integrato con Z Operations Unite per fornire il rilevamento e la risoluzione degli incidenti basati su chat AI utilizzando dati di sistema in tempo reale.

*"Nel nostro settore stiamo rapidamente imparando che l'AI è preziosa tanto quanto l'infrastruttura su cui girà, ha affermato **Ross Mauri, direttore generale di IBM Z e LinuxONE, IBM** "Con z17, portiamo l'AI al centro dell'azienda con il software, la potenza di elaborazione e lo storage necessari per rendere l'AI operativa rapidamente. Inoltre le aziende possono fornire all'AI i propri dati aziendali non ancora utilizzati in modo sicuro e conveniente".*

Integrazione completa tra hardware e software

IBM z17 è un sistema progettato per integrarsi completamente in ambienti ibridi, unendo innovazioni hardware, capacità software per AI e un ampio supporto per strumenti standard e aperti. Ciò consente di differenziare prestazioni e affidabilità, reinventando al contempo il modo in cui sviluppatori e operatori di sistema interagiscono e gestiscono IBM Z, tra cui:

Sistema operativo per l'AI - IBM ha presentato in anteprima anche z/OS 3.2, l'ultima versione del sistema operativo di punta per IBM Z, la cui disponibilità è prevista per il terzo trimestre del 2025. z/OS 3.2 è progettato per beneficiare delle funzionalità di AI accelerate dell'hardware e degli insight operativi basati su AI per la gestione del sistema stesso. Inoltre, z/OS 3.2 fornirà supporto per moderni metodi di accesso ai dati, database NoSQL ed elaborazione dei dati in cloud ibrido. Queste nuove funzionalità consentiranno al software di intelligenza artificiale di attingere a un set più ampio di dati aziendali e di ricavarne informazioni predittive sul business.

Operazioni IT unificate – oggi è stato annunciato anche IBM Z Operations Unite, che riunisce dati operativi provenienti da più fonti su IBM Z, in formato OpenTelemetry, per semplificare le operazioni di IBM Z con l'intelligenza artificiale. La nuova soluzione è stata sviluppata per accelerare il tempo di rilevamento delle anomalie, isolare l'impatto di potenziali incidenti e aiutare a ridurre il tempo di risoluzione. In sinergia con IBM Concert gli operatori di sistema possono trarre vantaggio dalla correlazione intelligente dei dati operativi. IBM Operations Unite sarà disponibile a partire da maggio 2025.

Acceleratore AI per l'efficienza di business – con le opzioni di espansione per alloggiare le schede IBM Spyre™ Accelerator, la cui disponibilità è prevista per il quarto trimestre 2025 tramite scheda PCIe, i clienti saranno anche in grado di eseguire un numero crescente di assistenti e agenti IBM, basati sui modelli Granite, in modo nativo su z17 senza assumersi il rischio aggiuntivo associato allo spostamento dei dati o della logica di business al di fuori della piattaforma. Insieme, queste soluzioni sono progettate per formare uno stack ottimizzato, che consente ai clienti di aumentare la produttività in modo sicuro e su larga scala.

Costruito per la resilienza: sicurezza e difesa informatica al centro del sistema.

IBM z17 rafforza la storia di una piattaforma che ha da sempre caratteristiche di sicurezza e resilienza indiscusse. I nuovi sviluppi basati sull'AI consentono l'implementazione di un'intelligenza aggiuntiva in quest'area sempre più importante per i clienti, poiché ogni giorno compaiono nuove minacce. Ciò include diverse nuove capacità, tra cui:

Gestione dei segreti: le soluzioni di HashiCorp, IBM Company, annunciate a marzo, sono ora disponibili su IBM Z per aiutare a standardizzare la gestione dei segreti nel cloud ibrido. IBM Vault utilizza la sicurezza identity-based per autenticare e autorizzare l'accesso a segreti, certificati, chiavi, token e altri dati sensibili. Grazie a Vault, i clienti possono avere un'unica soluzione per proteggere i carichi di lavoro critici gestendo in modo sicuro l'intero ciclo di vita dei segreti trasversalmente al dominio IT.

Sicurezza dei dati utilizzando l'intelligenza artificiale – sono fornite nuove capacità per scoprire e classificare i dati sensibili sulla piattaforma. Questa funzionalità è legata a Telum® II e usa l'elaborazione del linguaggio naturale e in modo che i dati mission-critical possano essere identificati e protetti. Inoltre, la nostra ultima soluzione di sicurezza basata sull'intelligenza artificiale, IBM Threat Detection for z/OS, è progettata per rilevare e identificare le anomalie potenzialmente dannose che potrebbero essere il risultato di un attacco informatico.

IBM estende il supporto AI-Enabled a IBM z17

L'esperienza di supporto completa e personalizzata di IBM aiuta i clienti IBM Z a soddisfare richieste che vanno oltre la manutenzione tradizionale. Fornito da IBM Technology Lifecycle Services, IBM Support for z17 aiuta i clienti a ottimizzare le prestazioni di picco, riducendo al minimo rischi e interruzioni delle operazioni mission-critical. Le funzionalità IBM per semplificare la correzione degli incidenti e contribuire a migliorare i tempi di risoluzione dei casi, basate su IBM watsonx, sono ora disponibili per i clienti IBM Z.

IBM offre storage sicuro e agile

IBM Storage DS8000 svolge un ruolo chiave come soluzione di storage integrata per IBM Z. L'ultima generazione di IBM Storage DS8000 ([10a generazione](#)) è progettata per sfruttare tutta la potenza di IBM z17, offrendo alle organizzazioni accesso a carichi di lavoro critici, prestazioni dei dati coerenti e ottimizzate e un'architettura modulare per adottare le più recenti tecnologie sostenute dalla ricerca IBM, in modo da alimentare la crescita del business monetizzando i dati. Insieme, IBM Z e IBM Storage offrono un'infrastruttura moderna, sicura e agile per carichi di lavoro critici.

Disponibilità

IBM z17 sarà disponibile al pubblico dal 18 giugno 2025.

Per ulteriori informazioni: ibm.com/it-it/products/z17

IBM Spyre™ Accelerator sarà disponibile a partire dal quarto trimestre del 2025.

Le dichiarazioni relative alla direzione e agli intenti futuri di IBM sono soggette a modifica o ritiro senza preavviso e sono solo indicativi di obiettivi e traguardi.

Informazioni su IBM

IBM è fornitore globale leader di cloud ibrido, intelligenza artificiale e competenze di consulenza. Aiuta i clienti in oltre 175 paesi a sfruttare le informazioni ricavate dai loro dati, ottimizzare i processi aziendali, ridurre i costi e ottenere un vantaggio competitivo nei loro settori. Più di 4.000 enti governativi e aziendali in aree infrastrutturali critiche come servizi finanziari, telecomunicazioni e sanità si affidano alla piattaforma cloud ibrida IBM e Red Hat OpenShift per realizzare le loro trasformazioni digitali in modo rapido, efficiente e sicuro. Le innovazioni di IBM nell'intelligenza artificiale, nell'informatica quantistica, nelle soluzioni cloud specifiche del settore e nella consulenza offrono opzioni aperte e flessibili ai nostri clienti. Tutto questo è supportato dall'impegno di lunga data di IBM per fiducia, trasparenza, responsabilità, inclusione e servizio.

Visita ibm.com/it-it per maggiori informazioni.

Per approfondire

- [z17](#)
- [Blog](#) Software
- [Blog](#) IBM Research

Contatti:

Titti Garau - *External Relations, IBM Italy*

email: titti_garau@it.ibm.com

mobile: 335 7248411

Disclaimer

¹. Dall'introduzione di IBM z16 nel 2022, i mainframe IBM Z sono stati in grado di supportare l'inferenza AI direttamente nel mainframe, rendendo possibile il punteggio del 100% delle transazioni in tempo reale anche in ambienti di produzione ad alto volume.

Report Celent: "[Mitigating Fraud in The AI Age](#)" di Neil Katkov, 04/08/2025, commissionato da IBM

². La differenza percentuale tra IBM z17, che elabora fino a 450 miliardi di operazioni di inferenza al giorno con un tempo di risposta di 1 ms utilizzando un modello di apprendimento profondo di rilevamento delle frodi sulle carte di credito e IBM z16 che può elaborare fino a 300 miliardi di richieste di inferenza al giorno con un tempo di risposta di 1 ms utilizzando un modello di rilevamento delle frodi sulle carte di credito.

Per IBM z17 il risultato delle prestazioni è estrapolato dai test interni IBM® in esecuzione su IBM Systems Hardware di tipo macchina 9175. Il benchmark è stato eseguito con 1 thread che eseguiva operazioni di inferenza locali utilizzando un modello di rilevamento delle frodi sulle carte di credito sintetico basato su LSTM (<https://github.com/IBM/ai-on-z-fraud-detection>) per sfruttare l'acceleratore integrato per AI. È stata utilizzata una dimensione batch di 160. Configurazione IBM Systems Hardware: 1 LPAR che esegue Red Hat® Enterprise Linux® 9.4 con 6 IFL (SMT), 128 GB di memoria. 1 LPAR con 2 CP, 4 zIIP e 256 GB di memoria che eseguono IBM z/OS® 3.1 con la funzionalità IBM z/OS Container Extensions (zCX). I risultati possono variare.

Per IBM z16, il risultato delle prestazioni è estrapolato dai test interni IBM che eseguono operazioni di inferenza locale in una LPAR IBM z16 con 48 IFL e 128 GB di memoria su Ubuntu 20.04 (modalità SMT) utilizzando un modello sintetico di rilevamento delle frodi con carta di credito (<https://github.com/IBM/ai-on-z-fraud-detection>) sfruttando IBM Integrated Accelerator for AI.

Il benchmark è stato eseguito con 8 thread paralleli, ciascuno bloccato sul primo core di un chip diverso. È stato utilizzato il comando lscpu per identificare la topologia core-chip. È stata utilizzata una dimensione batch di 128 operazioni di inferenza. I risultati sono stati riprodotti anche utilizzando una LPAR z/OS V2R4 con 24 CP e 256 GB di memoria su IBM z16. È stato utilizzato lo stesso modello di rilevamento delle frodi con carta di credito. Il benchmark è stato eseguito con un singolo thread che eseguiva operazioni di inferenza. È stata utilizzata una dimensione batch di 128 operazioni di inferenza. I risultati possono variare.
