IBM z17 makes more possible

Published: 8 April 2025 AD25-0015 Hardware

Overview

Today IBM® is announcing the IBM z17™, a major evolution of the iconic, fully integrated system with new AI capabilities across hardware, software and systems operations. Powered by the new IBM Telum® II™ processor, z17 expands the system's capabilities beyond transactional AI capabilities, enabling the platform to take on new workloads without additional infrastructure expenses. The AI+ era is enabling new opportunities, competitive advantage and growth; however, it can place significant demands on your IT. IBM z17 is designed collaboratively with clients to enable AI where it matters most to drive efficiency, innovation, and better business outcomes. Leverage AI to not only grow your business but to optimize and secure your infrastructure.

The Telum II processor features key innovations to meet the demands of today's workloads. It boasts a completely new data processing unit (DPU) which is engineered to accelerate complex input/output (I/O) protocols for networking and storage on the mainframe. The Telum II processor also features the 2nd generation of the on-chip AI acceleration, allowing AI inferencing to happen as close to transactions as possible for optimized response performance. In this generation, more types of AI acceleration are achieved with embedded compute primitives for encoder large language models (LLMs) as well as expanded quantization and matrix operations for efficiency. In addition, new intelligent routing capabilities enable applications to access all the AI accelerators across the processor drawer, even those not on the same physical chip. With collaboration from clients, IBM Z®has redefined AI at scale with innovations to help clients stay competitive in the market. Keeping client success in the forefront, IBM z17 is designed, for example, to process up to 5 million inference operations per second with less than 1 ms response time[1] . This allows enterprises to evaluate 100% of their transactions in real time reducing fraud exposure risk[2].

IBM z17 is engineered to leverage AI not only to grow your business but also to optimize and secure your infrastructure. Upon the availability of the IBM Z Spyre™Accelerator in 4Q 2025, IBM Z aims to transform the user experience on the platform by running generative AI technologies like IBM watsonx Code Assistant for Z and IBM watsonx™ Assistant for Z, fully on prem, to make informed decisions and perform tasks with confidence. Underpinning all of this is a strong legacy of security. IBM z17 helps clients enhance their security posture with AI-powered security

features and consolidated dashboards designed to simplify security and compliance management.

Overall IBM z17 is designed with a full stack approach with clients and the full ecosystem in mind. The value the innovations bring will drive business growth into the future.

1 DISCLAIMER: 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 (CCFD) model (https://github.com/IBM/ai-on-z-fraud-detection) to exploit the IBM 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

2 CELENT: Mitigating fraud in the AI age

Fueling Innovation and Growth with AI

IBM z17 makes more possible, enabling AI where it matters most to drive efficiency, innovation and better business outcomes. It was fully engineered from the chip through the software stack for AI, to optimize mission-critical transaction processing and data-intensive workloads. With this full stack optimization, enhanced AI acceleration capabilities, along with improved I/O density and efficient processor power management, IBM z17 delivers scalable AI inferencing in an energy optimized compute environment.

Real-time in-transaction insights on IBM z17

IBM z17 inferencing, powered by the Telum II processor, enables new capabilities for predictive AI, including:
- Enhanced compute primitives for encoder Large Language Models (LLMs)
- Model quantization support
- Enhanced matrix operations support

The new enhanced compute primitives are designed to extract insights and generate embeddings from a broader array of both structured and unstructured data. With the new embedded application support, LLMs can be leveraged in multiple ways—whether invoked directly from applications, exposed as REST API calls, or integrated within environments like IMS™, CICS®, and others. In addition, with the new support for model quantization and enhanced matrix operations, along with processor memory cache enhancements, IBM z17 is designed for optimized latency and throughput. Overall, these innovations can be combined to leverage multiple model strategy for optimized AI predictive scoring. The collective strength of these innovations has been demonstrated in our labs, showing that IBM z17 can process up to 450 billion inference operations per day with 1 ms response time[3].

To exploit hardware and system level acceleration capabilities, there are a multitude of software entry points for AI enablement. There are solutions directed at both z/OS and Linux on Z. Two key entry points are Machine Learning for IBM z/OS and a collection of AI frameworks and tools in AI Toolkit for IBM Z. Throughout experiences designing AI solutions with clients from the silicon up, it is clear that many clients have unique scenarios with a wide range of business challenges. IBM and our ecosystem partners continuously deliver new features and functions which can be found on our product pages.

Machine Learning for IBM z/OS (MLz) - MLz will leverage the acceleration capabilities of Telum II and IBM Spyre Accelerator (see SOD for availability), opening new opportunities to deploy AI models for advanced inferencing at scale. With z17, MLz will enable clients to leverage encoder LLMs (like BERT) that excel at understanding natural language and processing unstructured text, focusing on extracting key information for improved scoring results (e.g., increased accuracy, reduced false positives, etc.) and address a wider range of advanced AI use cases like fraud detection, anti-money laundering (AML), insurance claims, etc. to achieve business insights at scale.

MLz and AI Toolkit for IBM Z & LinuxONE both include the latest version of the IBM Z Deep Learning Compiler which allows the exploitation of Telum II capabilities. In addition, enhancements have also been made to IBM Z Accelerated for PyTorch, IBM Z Accelerated for TensorFlow, IBM Z Accelerated Serving for TensorFlow, and IBM Z Accelerated for NVIDIA Triton Inference Server, optimized to take full advantage of Telum II.

The IBM Telum II processor together with the IBM Spyre Accelerator (see SOD for availability) on IBM z17 open a world of possibilities with AI. IBM z17 with the Spyre accelerator will enable generative AI use cases on premises. These can range from running chatbots, assistants and agents to a variety of clients specific use cases for things like document processing, information search and data extraction, all while keeping their data secured and in place.

3 DISCLAIMER: 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.

Automating and transforming for efficiency

Digital transformation and IT efficiency are crucial for organizations to stay competitive, reduce operational costs, and accelerate innovation. Streamlining development and operations through modern technologies enables faster decision-making, improved scalability, and enhanced security, ultimately driving better client experiences, increased profitability, and long-term business success. Leveraging IBM

z17 with its engineered AI stack, businesses can achieve these goals, ultimately delivering more value through long-standing investments that are driving the core of the business. Paired with generative AI assistants, available on z17 with IBM Spyre Accelerator in 4Q25 (see SOD for availability), and strategic Transaction Processing Platform software, clients will have a modern framework to help transform the mainframe user experience with AI assistants and Agentic AI designed to boost productivity, improve skills and drive further simplification. z17 empowers clients to accelerate modernization with AI, to streamline IT operations with AI, and to boost productivity.

Accelerate the mainframe application lifecycle with generative AI and automation

IBM z17 will help assist in accelerating mainframe application modernization using generative AI. IBM watsonx Code Assistant for Z is a strategic component to this strategy. IBM watsonx Code Assistant for Z is designed to accelerate mainframe application modernization with automation and generative AI specially trained, tuned, and customized around IBM Z code and applications. This solution will be enabled on z17 with IBM Spyre Accelerator once delivered in 4Q, 2025. This solution addresses the end-to-end application lifecycle with capabilities including Understand, Refactor, Explain, Optimize, Transform, and Validate. It enables enterprises to adopt an incremental approach to mainframe application modernization, improving agility and productivity while mitigating risk.

IBM watsonx Code Assistant for Z continues to advance its capabilities and empower developers in successfully modernizing mainframe applications.

New enhancements include:
- Chat-style explanations, leveraging AI agents: Reduces skill gap in developers by enabling them to get context aware insights of code
- PL/I support for Code Explanation and Refactoring: Improve code understanding and business agility for PL/I applications
- Code Optimization support for COBOL: Improve application performance through prioritized recommendations and insights

In 2025, enterprises will be able to take advantage of new features, including, but not limited to: code generation, a simplified experience for application understanding, and additional language support for Code Explanation and Transformation.

In addition to IBM watsonx Code Assistant for Z, developers require effective and agile code testing to create, run, and enhance unit, early integration and functional tests for Mainframe programs without the need for data and middleware. IBM Test Accelerator for Z helps developers achieve these goals. As a first step, the next release will integrate Test Accelerator for Z and watsonx Assistant for Z, allowing developers to automatically generate functional tests using natural language prompts. With the introduction of Agentic AI in the upcoming release, the developer experience for test case creation, management, and the testing process itself will undergo a significant transformation. Additionally, by using Test Accelerator for Z alongside Test and Stress Testing (TST), developers can reduce costs related to hardware capacity.

For instance, TST can be used for just 20 days per year, while maintaining 24/7/365 test capacity through an 'always on' subscription model.

Streamline IT operations with AI Insights

A new set of products, IBM Z Operations Unite, IBM Concert®, and IBM Z Intellimagic Vision for z/OS, aiming to streamline IT operations with AI, will help clients drive new operational efficiencies with z17 in their data centers.

IBM Z Operations Unite is a key solution to this strategy, IBM Z Operations Unite enables client infrastructure teams to accelerate the resolution of operational issues and optimize systems management. With an observability framework at its core, IBM Z Operations Unite empowers users with continuous insights, seamless integration to existing tools, reduction of distracting alerts, and provides AI-generated recommendations to significantly boost operational efficiency and system availability. With IBM Z Operations Unite, users can streamline operational workflow and move faster to identify, isolate, investigate, and resolve problems that are impacting infrastructure health.

IBM Concert is another new integration to IBM Z that gives z/OS System Programmers the ability to assess their enterprise maintenance posture. With this information they can easily form resilient processes and avoid outages due to unapplied APARs. Concert collects and combines information such as APAR status from z/OSMF, software usage from IBM Z Software Asset Management, and manually defined SLA parameters to calculate a maintenance risk score. Concert allows users to visualize maintenance risk, prerequisites and dependencies so operations teams can clearly define their software upgrade roadmap. Concert users can also auto-initiate change tickets to kickoff APAR installation with the proper IBM Z SME.

Finally, IBM Z IntelliMagic Vision for z/OS is a cloud-based solution that modernizes mainframe performance analysis with proactive health insights spanning the end-to-end z/OS infrastructure. With advanced anomaly detection capabilities, dynamically customizable reports, context sensitive drill downs, and a modern and interactive GUI, IBM Z IntelliMagic Vision for z/OS augments the capabilities of both new and existing staff to more effectively and efficiently ensure application service levels and optimize overall performance health. Also, new reports will be available to highlight areas where z17 provides increased capacity capabilities and to highlight workloads that will benefit from the increased capabilities. Reports will be available that help clients specifically manage performance of Z workloads during the upgrade process to z17 and continuously once their z17 environments are operational.

Boost productivity and automate tasks

IBM z17 is designed to aid organizations in boosting productivity using generative AI, industry best-practices, and automating tasks within their business-critical mainframe environments and core applications. With rich support for open standards and practices, z17 is designed to enable seamless interoperability within a hybrid IT environment, ensuring flexibility and scalability. Secure data access empowers businesses to make informed decisions by leveraging real-time, accurate insights

from their core business and application data. Combining open practices and data helps unlock the full potential of hybrid cloud integration. z17 provides an agile way to manage its workloads spanning on-premises and cloud environments where mainframe services are needed across the enterprise. Together, these elements enable businesses to streamline operations, enhance collaboration, automate routine tasks, and drive innovation, driving toward increased efficiency and improved business outcomes.

One of the largest pain-points for mainframe clients is maintaining a skilled workforce to operate mainframe infrastructure while also finding ways to simplify their mainframe practices to deliver efficiency. IBM watsonx Assistant for Z is a revolutionary new approach for mainframe productivity with generative AI assistants, available on z17 with IBM Spyre Accelerator in 4Q25 (see SOD for availability). IBM watsonx Assistant for Z brings decades of experience and best practices to your Z users' fingertips and transforms how they engage with and manage the mainframe. This AI assistant provides users with accurate, up-to-date answers to their IBM Z queries and simplifies the execution of both complex and repeated tasks, by codifying the knowledge of Z experts into a trusted set of automations, that can be used by IBM Z users of all experience levels. Users also customize this AI assistant by ingesting your documentation for personalized experience. IBM watsonx Assistant for Z is designed to improve IBM Z users' autonomy and confidence by reducing the time to onboard new-to-Z users and accelerating knowledge transfer from Z experts.

To accelerate and simplify the deployment of IBM watsonx Code Assistant for Z and IBM watsonx Assistant for Z, clients can leverage the IBM AI Optimizer (announced on April 8, 2025), a fit-for-purpose software program designed to enable IBM Z clients to simply and quickly adopt generative AI workloads. IBM AI Optimizer for Z 1.1 Essentials Edition provides the following capabilities:

- Consolidated and automated installation and configuration of the following IBM programs:
  - IBM watsonx Code Assistant for Z
  - IBM watsonx Assistant for Z
- Automation, configuration, and validation of inferencing infrastructure required for IBM watsonx Code Assistant for Z and IBM watsonx Assistant for Z.
- Product health monitoring dashboard.

Beyond assistants, data access, open standards, open tooling, and open platforms are key to supporting increased productivity with IBM z17. For mainframe systems, this means running many of these tools natively on z/OS where the transactions and data reside. Data access to IBM z17 is vital to driving productivity and helping to minimize risk in an AI strategy, Db2® Analytics Accelerator for z/OS and IBM Data Gate® for watsonx are two key products supporting this journey. First, Db2 Analytics Accelerator provides unprecedented analytic processing performance for real-time query results from IBM Z data in place and is designed to benefit from enhancements in the IBM z17 infrastructure ranging from the increased cache to the new features of the Network Express adapters.

IBM Data Gate for watsonx is another key component in supporting client's existing AI strategy. It offers an efficient and simple approach to synchronize IBM Z data to the watsonx.data platform. The watsonx.data platform supports data transformation for analytics and AI model building off IBM z17 where existing AI investment and processing for training are used in the enterprise. It is designed to support synchronization of IBM Z data sources with external platforms.

In addition to data access and management, IBM Z Platform for Apache Spark is built specifically to accelerate the time to insight from analytics by providing a full framework of capability for data science natively on z/OS.

Another essential component to a productivity and efficiency strategy is ensuring modernization across the data center with a consistent container and DevOps experience. Red Hat OpenShift® Container Platform is a hybrid cloud application platform for providing a comprehensive, Kubernetes-powered application platform that simplifies deployment, management, and scaling of applications across various environments, offering flexibility, consistency, and enhanced security, while also supporting both cloud-native and traditional workloads. With IBM z17 system and Red Hat OpenShift there are enhanced capabilities to scale infrastructure, protect data at rest, in transit and in use, and to streamline infrastructure management. Embracing Red Hat OpenShift along with automation frameworks such as Terraform, an open-source, Infrastructure-as-Code (IaC) tool, and Red Hat Ansible® for provisioning, configuration management, and application deployment, can assist businesses with defining their environments in standard formats, rather than using manual processes or complex interfaces. The IBM z17 stack is engineered and optimized to embrace and support these practices across the enterprise for success today and provides a foundation for future enhancements.

In the age of AI where clients are looking to drive growth and efficiency, pricing becomes very important. With IBM z17, there are additional opportunities to evolve how clients consume and use their z17 capacity. Clients are empowered to unlock these new opportunities as they design, deploy, and scale new workloads on z/OS with optimized total cost of ownership and the agility of cloud-like pricing, exclusively with Tailored Fit Pricing for IBM z17.

- New technology on z17 enables workload-optimized pricing
- Unlock a pool of always-on subscription zIIP capacity for AI, OpenShift, and zCX workloads
- Consumption-based pricing delivers greater agility, so clients can meet the dynamic demands of new workloads

Reduce costs with efficient energy management and intuitive tooling

An IT solution that is as sustainable as possible across its lifecycle is key to meeting business goals and global environmental regulatory requirements. Partnering with IBM, which has decades of leadership in environmental sustainability, ensures that the environmental footprint of your IT infrastructure is in good hands.

IBM z17 is designed to provide these direct benefits for efficiency and cost effectiveness:

- **Run AI in a secure and energy-efficient environment.**Clients need to deploy AI for mission critical workloads without negatively impacting the energy consumption of their data centers. With IBM z17's full stack optimization across the Telum II I/O Data Processing Unit (DPU), the integrated AI accelerator, and Spyre Accelerator (see SOD for availability), along with improved I/O density and efficient processor power management, clients get all the AI inferencing they need without increasing their carbon footprint or energy costs.
- **Make sustainability more achievable and straightforward.** With IBM z17 at the heart of a client's infrastructure, they see immediate savings in energy consumption and carbon footprint, especially when compared to previous generation systems. The anticipated large IBM z17 system can reduce weight by 26%, data center floor space by 25%, and is estimated to reduce system power consumption by approximately 27%, compared to a similarly configured large IBM z16® system[4]. Enhanced tooling lets clients estimate and monitor savings using a customized product carbon footprint calculator and provides the ability to set up notifications for when energy consumption thresholds are exceeded. New for IBM z17 is the ability to track system and partition power consumption through the operating system, which clients can then use to calculate even more granular power metrics. These capabilities enable organizations to access, track, and report the exact level of data needed to prove continuing sustainability throughout the product lifecycle.

IBM z17 is designed for end-to-end sustainability, which reduces total cost of ownership and energy costs and helps businesses adhere to global environmental regulatory requirements and meet sustainability goals.

SRB Upgrade will not be supported on IBM z17

SRB Upgrade will not be orderable on z17, and SRB Upgrade cannot be carried forward to z17 from previous machine generations. Please note that this *only* affects the SRB Upgrade offering.

- SRB base capabilities (Shutdown Boost, Startup Boost, Recovery Process Boosts) will be supported on z17.
- There will be no changes to existing support/availability of:
  - SRB base capabilities on IBM z15® and z16.
  - SRB Upgrade support/availability on z15 and z16.

4 DISCLAIMER: Based on an expected large IBM z17 system configuration based on an actual historical large IBM z16 system configuration. IBM z17 is Max208 with 23TB memory, 56 active processors, 3 IBM Virtual Flash Memory, 14 ICA-SR 2.0, 7 PCIe+ I/O drawers with 69 FICON Express32 – 4P LX, 12 OSA-Express7S 1.2 GbE SX, 18 Network Express LR 10G, and 4 Crypto Express 8S (2 HSMs). The IBM z16 is configured to provide the same hardware capability. Power consumption is based on the Power Estimation Tool. Results may vary.

IBM Resource Link®: Power and weight estimation.

This IBM Resource Link web page is the source for Power and weight estimation.

Securing the most important data

IBM Z is an industry leader in cyber security and resilience with 8x9s of availability[5], flexible capacity options for adding compute within the system, the ability to swap full environments, and the industry's first quantum-safe system[6]. IBM z17 delivers capabilities designed to protect mission critical data by leveraging quantum-safe cryptography, artificial intelligence, and automation. Additionally, IBM z17 will enable clients to simplify security management and uncover insights for enhanced visibility into security posture and compliance. IBM z17 leverages hardware, firmware, and software to bring a comprehensive suite of security capabilities to maintain critical business operations.

Employing the 2024 NIST standards for quantum safe cryptography for IBM Z

The Crypto Express 8S is the latest generation IBM hardware security module (HSM). It performs top-level security processing and high-speed cryptographic functions with a high throughput rate that reduces latency. The HSM supports two cryptographic APIs: CCA and EP11.

The CCA API introduces support for NIST algorithms to meet most of the PQC / Quantum Safe public key use cases and key management support for the PQC keys, allowing integration with enterprise key management offerings and services such as IBM Unified Key Orchestrator for z/OS (UKO for z/OS) and IBM Enterprise Key Management Foundation (EKMF), respectively. The added PQC algorithms are:

- ML-KEM (NIST Federal Information Processing Standards 203 (FIPS 203)), a Module-Lattice-Based Key-Encapsulation Mechanism Standard for general encryption.
- ML-DSA (NIST FIPS 204), a Module-Lattice-Based Digital Signature Standard for authentication of digital signatures, including HashML-DSA.

Other CCA updates include:

- Technical Report TR-31 (X9.143) native support in all operational APIs. TR-31 is the standard key block in the financial services industry. Users of the CEX8S are now able to create TR-31 keys under the CCA Master Key (CCA-MK) like any other key type, and use them in CCA services, not just for key transport.
- Secure Hash Algorithm 3 (SHA3) is now available for hash services and digital signature generation/verification. The CCA host library for Linux® on IBM Z now also has support to send SHA3 requests to the Central Processor Assist for Cryptographic Functions (CPACF).
- RSA updates include RSA 8192 bit key support, Optimal Asymmetric Encryption Padding (OAEP) 2.1 encoding for RSA encryption, decryption, and import / export of RSA-AESKW wrapped keys - allowing key exchange between CCA and PKCS#11 APIs popular at major cloud cryptography service providers. (RSA is the public-key cryptosystem invented by Rivest, Shamir, and Adleman; AESKW is the Advanced Encryption Standard Key Wrap.)

The EP11 API provides:

- An additional FIPS 140-2 approved image. clients can switch from the regular to the FIPS approved image on a per AP basis if required by regulators. Usage of the regular image is preferred as it contains the latest features and bug fixes. The FIPS image will match the version as mentioned by the most recent Cryptographic Module Validation Program (CMVP) certificate available for the module.
- An update that adds support for the BLS digital signature algorithm with the BLS12-381 pairing-friendly curve to EP11. BLS signatures are deterministic and non-malleable and used by the Ethereum 2 crypto currency system. It allows for both aggregation of signatures and public keys. Additionally, support for the key derivation required by deterministic hierarchical wallets as defined in EIP2333 is added (BLS is a digital signature scheme invented by Boneh, Lynn, and Shacham).
- An update that adds support for assigning Crypto Adapters in EP11 mode to guest configurations running in secure execution mode. Adapters must be bound and associated with the guest configuration to become operational. Association employs a guest secret known to the owner of the image only to ensure generated secrets can be used by this secure guest image only, even if the adapter gets moved to a different configuration without zeroization.

The EP11 will also provide the ML-KEM and ML-DSA vendor mechanisms, detailed above, later this year.

Simplified Quantum-safe security journey for IBM Z

IBM Z Crypto Discovery & Inventory v 1.1.0 solution will aid clients in monitoring, uncovering and prioritizing cryptographically relevant security vulnerabilities. Data security is the foundation of every business transaction. Enterprises have relied on traditional cryptography and encryption methods to safeguard their data, applications and endpoints for decades. Traditional encryption algorithms that would take centuries to break by using conventional computers may be broken within hours with a cryptographically relevant quantum computer.

As quantum computing advances, sensitive information secured by today's encryption standards and practices may potentially be left vulnerable. To address this challenge, organizations need to migrate to quantum safe cryptography, which will be a multiyear effort. So, it is essential that organizations start planning their journey now! One of the first steps on the quantum-safe journey will require organizations to understand where cryptography is being used in their environment and build a cryptographic inventory; a holistic, reuseable asset that can help uncover potential vulnerabilities and help with creating a risk-based prioritized migration roadmap.

IBM Z Crypto Discovery & Inventory will simplify the process of how clients discover uses of cryptography within their Z infrastructure by analyzing and consolidating crypto relevant statistics from several data sources. Clients can leverage IBM Z Crypto Discovery & Inventory to help build an inventory to understand their cryptographic landscape, assess risks, manage life cycles, plan for post quantum migration and more effectively respond to security incidents and changing cryptography standards.

IBM Z Security and Compliance Center is an automated compliance solution which gathers data from IBM Software and Products on IBM Z and IBM® LinuxONE platforms to simply audits, saving time and efforts. IBM Z Security and Compliance Center is now enhanced to include a security framework that will enable clients to simplify security management and uncover insights for enhanced visibility into their security posture, for instance, the IBM Z Crypto Discovery & Inventory can be used through this framework. IBM Z Security and Compliance Center now integrates with IBM Concert enabling complete enterprise coverage for compliance. Additionally, IBM has added new capabilities such as Security Patch Insights for IBM z/OS and Custom Goals.

Cyber Resiliency Solutions

IBM Z Flexible Capacity for Cyber Resiliency clients can now perform Infrastructure Tests to validate that processes, connections, and automation are in place. For example: SAN, network, distributed systems connected to the IBM Z machine, 3rd party license keys, lights-on, GDPS® scenarios, etc. Available exclusively on IBM z17, the Infrastructure Test capability delivers increased flexibility to validate disaster recovery-readiness and help manage compliance requirements.

- Conduct Infrastructure Tests using an *isolated copy of production workload*
- Up to 4 tests per year, up to 10 days per test
- Tests can be ordered via pre-paid Feature Code 0824

The latest release will provide integration with IBM Threat Detection for z/OS, enhanced validation capability, and support for clients using Copy Services Manager (CSM) for their High Available / Disaster Recovery. In addition, there are two functionalities related to the anti-malware topic of the SOD found here

IBM Vault Self-Managed

IBM Vault Self-Managed extends the power of HashiCorp Vault to the mainframe environment. This solution provides a standardized enterprise secrets management solution that builds on the security, resiliency and scalability of IBM Z and IBM® LinuxONE.

For more information on options and ordering, please read the recent announcement.

Simplify Compliance and Improve Productivity

Building on the critical hardware security module management functionality which version 10.0 provides, the Trusted Key Entry (TKE) 10.1 delivers additional support and improved user experience enhancements.

The TKE 10.1 offers:

- New wizard functionality which streamlines the process of creating new CCA master key parts and streamlines loading CCA master keys using key parts found on smart cards.
- A new version of the smart card utility program utility simplifies the experience of initializing and managing the lifecycle of every type of smart card used on the TKE.

– And a new dark blue smart card, part 03GN391.
Additionally, the new version adds support for creating TR-31 operational key parts and loading TR-31 operational keys into HSMs from the TKE. The new dark blue smart card is required if the key parts will be kept on smart cards.

For more information about all these features, see the Cryptographic Services ICSF Trusted Key Entry Workstation User's Guide.

5 CLAIM FOR z/OS: For clients running z/OS v3.1 or higher with a configured high availability IBM software stack on IBM z16 or IBM z17, users can expect up to 99.999999% availability or 315.58 milliseconds of downtime per year when using a GDPS 4.7 Continuous Availability (CA) configuration and workloads. DISCLAIMER: The claim is based on IBM internal data and a GDPS CA three-site configuration, 2 active Sysplex sites and 1 Disaster Recovery (DR) site, consisting of z/OS 3.1 or higher with a Recovery Time objective (RTO) of 2 minutes or less, one of the required GDPS CA IBM middleware stack workloads and replication products running on IBM z16 or IBM z17. GDPS CA includes resiliency features such as Parallel Sysplex® enabled data sharing applications, GDPS Metro Mirror replication (Hyperswap®), software replication, and other CA configuration documented high availability features. A supported GDPS CA middleware stack could include CICS v6.2, IMS v15.5, MQ® v9.4, and Db2® v13 or at later releases. Clients must follow maintenance, configuration, capacity planning and testing best practices for the entire software stack and hardware configuration. This includes enabling all the resiliency technology for their workloads as defined by GDPS CA, z/OS,56© 2025 IBM Corporation 56 and workload related software products. Other configurations may have different availability characteristics.

6 CLAIM: IBM z16 is the industry's first quantum-safe system. DISCLAIMER: IBM z16 with the Crypto Express 8S card provides quantum-safe APIs providing access to quantum-safe algorithms which have been selected by NIST to become part of its post-quantum cryptographic standard. NIST Announces First Four Quantum-Resistant Cryptographic Algorithms. Quantum-safe cryptography refers to efforts to identify algorithms that are resistant to attacks by both classical and quantum computers, to keep information assets secure even after a large-scale quantum computer has been built. Source: Quantum Safe Cryptography (QSC). These algorithms are used to help ensure the integrity of a number of the firmware and boot processes. IBM z16 is the Industry-first system protected by quantum-safe technology across multiple layers of firmware. According to Peter Rutten, Research Vice- President IDC, "z16 is the industry's first quantum-safe computing platform."

Planned availability date

June 18, 2025

New-build systems:
– IBM z17 Model ME1
– Features and functions for IBM z17

Upgrades available:

– IBM z15 Model T01 upgrades to IBM z17 Model ME1
– IBM z16 Model A01 upgrades to IBM z17 Model ME1

MES orders for IBM z16 that include the following features

Field-installed features and conversions on IBM z17 that are delivered solely through a modification to the machine's Licensed Internal Code (LIC).

– HMA IBM Z Hardware Management Appliance on IBM z17
– TKE IBM Z Management on IBM z17
– Field-installed Frame Rolls from IBM z15 to IBM z17 or from IBM z16 to IBM z17

Support for all remaining MES orders for IBM z17 Model ME1 offerings:

September 18, 2025

The offerings in this announcement might not be available for purchase in all countries or regions. Purchase availability can be affected by multiple factors that include support availability, service availability, and government regulations. Contact your IBM representative or IBM Business Partner for availability information.

Supporting information

– IBM Z IntelliMagic Vision for z/OS
– IBM Z and IBM® LinuxONE Container Registry
– Python AI Toolkit for IBM z/OS
– IBM Z Platform for Apache Spark 1.1.0
– IBM Z Connectivity Handbook
– IBM Concert
– IBM watsonx Code Assistant for Z
– Overview of Data Gate for watsonx
– Db2 Analytics Accelerator on Z
– Red Hat OpenShift on IBM Z and IBM® LinuxONE
– Quantum-safe security for IBM Z
– IBM Hyper Protect product family
– IBM Secure Execution for Linux on IBM Z and IBM® LinuxONE
– IBM Synthetic Data Sets Announcement

Other related announcements

Preview: z/OS 3.2, unlocking the value of the IBM z17 IBM plans to release z/OS 3.2, the next release of its flagship operating system for IBM Z, in the third quarter of 2025. It is planned to support the IBM z17 and bring AI-infused capabilities, improved security, and simplified management to IBM Z, while fueling innovation, growth, andoperational efficiency.

IBM Application Delivery Foundation for z/OS

IBM intends to deliver anti-malware for IBM z/OS IBM plans to provide a software solution that introduces cyber anomaly detection and notification for the z/OS platform to mitigate the potential risk of malicious software. IBM plans to provide the option of quarantine functionality that further extends existing remediation options. It is the intent for these combined functions, per NIST guidelines, to be used by the client to satisfy compliance regulations requiring anti-malware coverage for z/OS. This intent includes standards such as the Payment Card Industry Data Security Standard (PCI DSS) version 4.0.

Enhancements to software pricing Technology Transition Offerings:

Technology Update Pricing for the IBM z17 extends the software price-performance provided by Advanced Workload License Charges (AWLC) and Country Multiplex License Charges (CMLC) for IBM z17. The Revised Transition Charges for Sysplexes or Multiplexes offerings provides a transition to Technology Update Pricing for the IBM z17 for clients who have not yet fully migrated to IBM z17. This ensures that aggregation benefits are maintained and phases in the benefits of Technology Update Pricing for the IBM z17 pricing as clients migrate.

For additional information about software pricing for the IBM z17, see Software Announcement AD25-0148, dated April 8, 2025.

Statement of general direction

IBM Spyre AI Accelerator: IBM intends to deliver the IBM Spyre AI Accelerator on Z, which is a PCIe-attached AI accelerator designed to support products, assistants, and services that leverage large language models on the IBM Z platform. The IBM Spyre AI Accelerator on IBM z17 is designed to bring generative AI to mainframes, empowering enterprises to scale-up AI capabilities and workloads. Clients will be able to deploy large language models (LLMs) including encoder, decoder, encoder-decoder architecture LLMs on the IBM Z platform, and benefit from the robust security, performance, and reliability it offers. IBM Z software products that will depend on Spyre Accelerators for full on-prem deployment include watsonx Code Assistant for IBM Z and watsonx Assistant for IBM Z. A hybrid deployment model is available for watsonx Assistant for IBM Z before Spyre Accelerator availability, as announced here.

Other AI on IBM Z products that are also planned for Spyre AI Accelerator exploitation upon availability include Machine Learning for IBM z/OS, and components of AI Toolkit for IBM Z & LinuxONE like IBM Z Accelerated for Nvidia Triton Inference Server and IBM Z Accelerated for PyTorch. IBM Spyre AI Accelerator is planned to be available starting in 4Q 2025, in accordance with applicable import/export regulations.

AI Toolkit for IBM Z & LinuxONE with Spyre: IBM intends to deliver capabilities to exploit IBM Z Spyre AI accelerator with components of the AI Toolkit for IBM Z & LinuxONE. AI Toolkit for IBM Z & LinuxONE intends to deliver support for encoder large language models (LLMs) like BERT (Bidirectional Encoder Representation from

Transformers) which clients will be able to infuse into their IBM Z applications to leverage AI at scale. Components of the AI Toolkit for IBM Z & LinuxONE that will exploit these capabilities using the Spyre AI Accelerator upon availability includes IBM Z Accelerated for Nvidia Triton Inference Server and IBM Z Accelerated for PyTorch.

Machine Learning for IBM z/OS with Spyre: IBM intends to deliver capabilities to exploit IBM Z Spyre AI accelerator with Machine Learning for IBM z/OS Enterprise Edition. Machine Learning for IBM z/OS intends to deliver support for encoder large language models (LLMs) like BERT (Bidirectional Encoder Representation from Transformers) which clients will be able to infuse into their IBM Z applications to leverage AI at scale. Machine learning for IBM z/OS will exploit these capabilities using the Spyre AI Accelerator upon availability.

Unlocking deeper insights for Db2 z/OS with AI on IBM Z: IBM intends to provide advanced vector database capabilities that exploit the IBM z17 hardware AI acceleration features with Db2 for z/OS. We're bringing together capabilities to enable clients to unlock deeper insights from their structured and unstructured data.

Changes to TKE: IBM z17 is the last machine family where the tower hardware server form factor of the Trusted Key Entry (TKE) will be supported. The 1U TKE is available on the z17, and the 1U TKE will be the only supported hardware in the future. For console room use of the 1U TKE, IBM recommends that there is consideration to mount the 1U TKEs in a mini rack with acoustical noise protection.

Removal of support for White Smart Cards: Due to changes in technology, IBM is withdrawing support for white smart cards (parts 45D3398, 74Y0551, and 00JA710). White smart cards will not work on any release of TKE past release TKE 10.1. To prepare for this, clients must purchase new smart cards through Feature Code 0889 or 0886, initialize a set of smart cards that corresponds to each of your existing white smart cards, and copy the content from each old to new smart card.

Future Direction for IBM Threat Detection for z/OS: IBM intends to add network anomaly detection to its IBM Threat Detection for z/OS product.

Future direction for tagging sensitive data on IBM Z: IBM intends to leverage the Telum II Processor, artificial intelligence, and machine learning technology to identify and classify sensitive data in z/OS data sets. This will provide an automatic function to discover, classify and tag data which will support organizations in implementing more targeted and effective security measures to protect that information.

Future direction for IBM Terraform - IBM intends in the near future to support IBM Terraform for Z and IBM® LinuxONE to enable the configuration and management of Z resources through Infrastructure as Code.

New support for BCPII: BCPii v1 and HMC/Support Element (SE) Simple Network Management Protocol (SNMP) are being deprecated. No future feature enhancements will be added to BCPii v1 or SNMP, but there are no plans to remove existing support

for BCPii v1 or HMC/SE SNMP. All future automation functional enhancements will only be done for BCPii HWIREST/v2 and HMC WebServices APIs.

Removal of Support for Constrained Transactional Execution: IBM z17 is planned to be the last IBM Z hardware generation to support Constrained Transactional Execution. On subsequent IBM Z hardware generations, the TBEGINC instruction that is used to start a Constrained Transaction will receive an operation exception, and no Constrained Transactional Execution will be initiated or supported.

IBM recommends that programs currently using Constrained Transactional execution should dual path their usage, so that when running on an IBM Z hardware generation that supports Constrained Transactional Execution, Constrained Transactional Execution continues to be used; but when running on an IBM Z hardware generation that does not support Constrained Transactional Execution, alternative mechanisms are used instead.

IBM z17 provides several new Perform Locked Operation (PLO) functions that are intended to provide a set of alternative capabilities that can be used in lieu of Constrained Transactional Execution. Note that many other existing serialization mechanisms can also be used to provide the necessary atomicity as an alternative to the use of Constrained Transactional Execution.

Note that support for Non-Constrained Transactional Execution is also reduced starting with z17. Starting with z17, the TBEGIN instruction used to start a Non-Constrained Transaction will always complete with with Condition Code 1 and with no Transactional Execution initiated, requiring the program to use its mandatory "fallback path" for execution. It is planned that this reduced support for Non-Constrained Transactional Execution will continue on subsequent IBM Z hardware generations.

Removal of Support for 10 Gb Coupling Express3 LR Coupling Links (CL5): z17 introduces new support for Coupling Express3 LR 25G Coupling Link adapters to provide long-distance coupling link connectivity with 25 gigabit bandwidth (CL6), in addition to supporting the existing 10 gigabit bandwidth using Coupling Express3 LR 10G (CL5).

The machine after z17 is planned to be the last IBM Z hardware generation which will support the 10 gigabit bandwidth (CL5) coupling links. The subsequent hardware generation after that is planned to support only 25 gigabit bandwidth CL6 coupling links for long-distance coupling link connectivity.

This approach supports IBM's usual n/n-2 sysplex connectivity policy regarding the transition off of 10 gigabit bandwidth CL5 coupling links and onto 25 gigabit bandwidth CL6 coupling links for long-distance connectivity. clients should plan to complete this transition within their sysplex environments before introducing a machine of the future IBM Z hardware generation that only supports 25 gigabit bandwidth CL6 coupling links for long-distance coupling link connectivity; all earlier machines that will need to connect to such a machine must also have already migrated to the use of 25 gigabit bandwith CL6 coupling links by that time.

T10 Data Integrity Field (T10 DIF): As per a March 2025 Statement of Direction announcement IBM z16 is intended to be the last IBM Z generation to support T10 Data Integrity Field (T10 DIF).

Future Linux Distribution Support: The following Linux distributions plan to support IBM z17 and IBM® LinuxONE Next generation models for a final time, and do not support future generation IBM Z and IBM® LinuxONE generation models.
– RHEL 8

The upcoming SUSE Linux Enterprise Server 16 will support IBM z17, and SUSE is actively collaborating with IBM and the communities to extend support for next generations of the IBM Z platform.
Red Hat and IBM intend to deliver Red Hat OpenShift AI on IBM Z and IBM® LinuxONE.

Statements regarding IBM plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM. Information regarding potential future products is intended to outline our general product direction, and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Technical information

Core System

IBM z17 is the 3rd generation of liquid-cooled systems built on the 19” frame concept. The system configurations will be like those of the z16 intelligent Power Distribution Unit (iPDU) offerings. The Liquid-Cooled CP (model ME1) system configurations running at 5.5GHz will support 1-4 Central Processor Complex (CPC) drawers and up to 12 I/O drawers, depending on machine type and model feature. Only iPDU Power distribution option will be supported.

The system will support a new CP chip, with two chips assembled into a Dual Chip Module (DCM) per socket, and four sockets per CPC drawer. Each CP chip will have an 8-core maximum core count, corresponding to a maximum of 16 cores on a DCM, incorporating functions like I/O accelerator cores and spare core management.

New Differential Dual In-Line Memory Modules (D-DIMM) supporting Double Data Rate 5 Dynamic Random-Access Memory (DDR5 DRAM) in size increments of 32, 64, 128, 256, and 512GB will be installed in 4 to 6 groups of 8. Memory configurations will include counts of 32, 40, and 48 D-DIMMs per drawer, with RAIM functionality included in the D-DIMM size quoted. The system will support up to 16 TB of client memory per CPC drawer.

The supported fanout slot maximum remains at 12 per CPC drawer. Each CPC drawer will support up to 6 I/O PCIe Hubs used to connect to I/O Drawers.

- The I/O PCIe hubs support Gen5 x16 into the fanout and will drive Gen4 x16 (Bifurcated to 2 at Gen4 x8) out of the fanout to the I/O Drawers.
- Short Range Coupling (ICA SR 2.0 card) Fanouts will support Gen3 x16 (Bifurcated to 2 at Gen3 x8).
- The form factor of the CPC drawer's I/O Fanout slot has changed, and therefore new I/O PCIe Hubs will be used for extension to I/O Drawers, and new ICA SR 2.0 coupling cards will be used to interface to other servers in a Sysplex.

The system I/O will be supported by the same PCIe+ I/O Drawer as in z16. A subset of I/O Cards will be carried forward from z15 and z16. The system I/O will be supported by the same PCIe+ I/O Drawer as in z16. A subset of I/O Cards will be carried forward from z15 and z16. FICON/FCP, OSA and Crypto functionality will be supported in the I/O drawer.

Next gen I/O infrastructure

IBM z17 will provide an I/O Data Processing Unit (DPU) on the mainframe processor chip that provides processing capability for networking (OSA) and storage (FICON®/FCP) protocols. This will enable IBM z17 to provide better I/O channel throughput along with improved I/O density.

Network Express adapter

The introduction of the Network Express adapter on IBM z17 will enable converging of legacy OSA and ROCE into one hardware offering. The OSH CHPID type will operate exclusively in Layer-2 mode but will support all legacy functions available with OSD*, utilizing enhanced QDIO (EQDIO) architecture for z/OS, z/VM®, and z/TPF operating systems. Using the new NETH Function ID (FID) type, Linux on Z and the Shared Memory Communications – Remote (SMC-R) protocol can simultaneously utilize the same Network Express ports using a native-PCIe architecture for Linux networking and RDMA capabilities.The new NETD FID type offers a physical function that unlocks additional card features, while the owning Linux LPAR (setup) handles management tasks. This setup enables a more streamlined and efficient management process.

*A z/VM VSwitch supporting Network Express OSH does not currently support z/OS guests exploiting an EQDIO uplink port. In the interim, clients will be required to use either a guest-attached OSH device or existing functionality available with OSA-Express7S adapters.

IBM zHyperLink™ Express 2

zHyperLink is a connectivity method that dramatically reduces latency by interconnecting the IBM Z system Central Electronics Complexes (CECs) directly to supporting disk storage controllers (e.g. select IBM DS8900s and DS8A000s). z17 will introduce the next generation zHyperLink Express card which enables PCIe Gen4 connectivity to supporting storage controllers.

FICON Express32-4P

FICON Express32-4P will provide the next generation 32G Fibre Channel adapter, supporting both FICON and FCP channel types. These 4 port adapters will provide increased port (physical channel) density when compared to the prior generation 2 port FICON express adapters. Authentication and encryption capabilities in support of IBM Fibre Channel Endpoint Security are also included.

For information regarding z17 qualified Switch, Director and DWDM environments, please consult the qualification letters linked below:

DWDM Qualification letters

Switch and Director Qualification letters

Coupling

Introduction of Coupling Express3 LR 25G for Long Reach Coupling

IBM z17 introduces Coupling Express3 LR 25G, a two-port 25 gigabit ethernet-basedlong reach coupling adapter that carries a new coupling channel type: CL6. Also available on z17 is Coupling Express3 LR 10G, a two-port 10 gigabit ethernet-based long reach coupling adapter that carries the legacy coupling channel type CL5. These adapters each reside in a PCIe I/O drawer card slot, support up to 4 CHPIDs per port, and support 32 or 8 subchannels (devices) per CHPID. The maximum number of Coupling Express3 LR adapter features is 32 per IBM z17 (64 ports). The adapters both require a 9u single-mode fiber cable and are designed to drive distances up to 10 km unrepeated and up to 100 km with a qualified DWDM.

Coupling Express3 LR adapters provide coupling and STP timing connectivity between servers. Coupling Express3 LR 25G only connects with another Coupling Express3 LR 25G on a z17. Coupling Express3 LR 10G is used to connect with Coupling Express2 LR on a z16 and Coupling Express LR on a z15. Though Coupling Express3 LR 10G may also be used to connect with another Coupling Express3 LR 10G on a z17, it is recommended to order Coupling Express3 LR 25G for z17-to-z17 connections to help ensure long reach coupling connectivity with future processor generations. Documentation in the PR/SM Planning Guide will document this new coupling long-reach adapter and channel type.

STP/Time Synchronization IBM Z and IBM® LinuxONE

- **Implementation/support of NTPv4/Chrony** Server Time Protocol (STP) on z17 will implement NTPv4/Chrony for its support of the Network Time Protocol (NTP). This will improve the overall resiliency of IBM Z's NTP implementation for STP. It also will improve the accuracy and stability capabilities of IBM Z STP by making use of the full suite of NTP algorithms via Chrony.
- **Implementation/support of the Network Time Security (NTS) protocol for NTP**z17 introduces support of NTS for NTP (NTS4NTP). NTS is an approved IETF standard (RFC 8915) that provides cryptographic security for the client-server mode of NTP. NTS uses Transport Layer Security (TLS) and Authenticated

Encryption with Associated Data (AEAD) to 1) allow NTP users to obtain time in an authenticated manner 2) ensure that time data received is authenticated and secure 3) protect against attacks such as spoofing, man-in-the-middle and replay.
- **Support of Mixed Mode of NTP and PTP** z17 STP introduces support of mixed mode of NTP and PTP. Mixed mode operation improves the overall STP resiliency capabilities by allowing STP to use up to five external time sources (3 NTP, 2 PTP) simultaneously. Mixed mode operation also improves STP accuracy and stability.

Removal of support for use of Virtual Flash Memory for Coupling Facility images

As per an October 2023 Statement of Direction z17 will not support the use of Virtual Flash Memory (also known as Storage Class Memory) for Coupling Facility images. Clients who were using Virtual Flash Memory in the Coupling Facility as an “overflow” mechanism for IBM MQ Shared Queue structures must switch to use other alternatives, such as using larger MQ Shared Queue structures, using MQ shared message data sets (SMDS), or using MQ techniques to offload MQ message data to Db2.

Removal of support for Coupling Facility image use of dedicated general-purpose processors

As per an October 2023 Statement of Direction z17 will not support the use of dedicated general-purpose processors (GPs) to provide processing capacity for Coupling Facility (CF) images. Clients who were using dedicated GPs for their Coupling Facility images must switch to one of the other supported processor configurations for CF images. Dedicated Integrated Coupling Facility (ICF) processors, shared ICF processors, and shared general-purpose processors remain supported for use by Coupling Facility images on z17.

Systems Management and infrastructure

The Hardware Management Console (HMC) 2.17.0 can only be run on the Hardware Management Appliance (HMA) on the z17, z16, or z15 HMA. That code level cannot run on a Standalone HMC. The HMC 2.17.0 highlighted enhancements include the following:

- **Dual Control** adds an extra layer of security for critical tasks on the HMC as well as can protect against user mistakes for those critical tasks. Dual control enabled tasks require another level of verification of the actual selections by a user from an approver before they can be run.
- **Remote Code Load** enhancements will provide alerts when remote code load is scheduled, running, and completed. IBM Resource Link will also execute health checks prior to or at scheduling time, and any issues will block remote code load scheduling rather than scheduling failing after further verification on the HMC.
- **Single Sign On (SSO)** will provide federated HMC user logon support using OpenID Connect (OIDC).
- **Manage System Time** enhancements will provide Network Time Security connections to Network Time Protocol (NTP) external time servers. A mix of NTP and Precision Time Protocol (PTP) servers will also be supported, and PTP communication support will now support Unicast in addition to Multicast.

- **Base Control Program internal interface (BCPii) HWIREST/v2** will provide an enhanced security control by mapping a z/OS user to an HMC user, and BCPii v2 requests will be limited to the user permissions of that HMC user. BCPii v2 will also provide asynchronous notification support for requests, and BCPii will now support requests sent to HMC targets enabling additional set of functionalities.
- **File system import/export** will enable an option for HMC tasks to import and export files directly to the remote browsing workstation without needing an FTP server or USB device.
- **User data replication** enhancements will now replicate HMC defined user patterns and templates to managed SEs. All HMC users and user data will now automatically be replicated to managed SEs, and clients can more easily use HMC defined users to manage their HMCs and SEs.

Performance

IBM's Large Systems Performance Reference (LSPR) method is designed to provide comprehensive IBM z/Architecture® processor capacity ratios for different configurations of central processors (CPs) across a wide variety of system control programs and workload environments. For IBM z17, the z/Architecture processor capacity indicator is defined with a 4XX, 5XX, 6XX, or 7XX notation, where XX is the number of installed CPs.

The LSPR applies uniformly to z/OS, z/VM and Linux on Z operating environments.

Single processor capacity of IBM z17 for equal n-way at common client configurations is approximately 11% greater than on IBM z16 with some variation based on workload and configuration[7]. The largest IBM z17 is expected to provide approximately 15% more capacity than the largest IBM z16 with some variation based on workload and configuration[8]. Within each single drawer, IBM z17 provides 20% greater capacity than IBM z16 for standard models and 15% greater capacity on the max config model, enabling efficient scaling of partitions[9]. The IFL and zIIP processors on the IBM z17 also provide an optional IBM z17 multi-threading technology capability; with the multi-threading function enabled, the performance capacity of a IFL or zIIP is expected to typically be up to 25% higher than without the multi-threading function enabled[10].

The LSPR contains the Internal Throughput Rate Ratios (ITRRs) for IBM z17 and the previous generation IBM Z processor families based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that a user may experience will vary depending on considerations such as the level of multiprogramming in the user's job stream, the I/O configuration, the workload processed, and the LPAR configuration. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance estimates stated. For more detailed performance information, consult the IBM Z Large Systems Performance Reference (LSPR) website.

7, 8, 10 Disclaimer: Based on internal measurements. Max config models IBM z17 Max208 and IBM z16 Max200 were used. Results may vary by client based on

individual workload, configuration and software levels. Visit LSPR website for more details. Visit the IBM Z LSPR website for more details.

9 Disclaimer: Based on internal measurements. Standard models include IBM z17 Max43, Max90, Max136 and Max183 and IBM z16 Max39, Max82, Max125 and Max168. Max config models are IBM z17 Max208 and IBM z16 Max200. Results may vary by client based on individual workload, configuration and software levels. Visit LSPR website for more details. Visit the IBM Z LSPR website for more details.

System reliability, availability and serviceability

The IBM z17 platform builds upon the reliability, availability, and resilience of predecessor IBM Z server platforms with a rich suite of improvements.

On the Telum II processor, there is improved fault tolerance for the IBM Z Integrated Accelerator for AI (AIU). Should an AIU fail, workload can be automatically re-routed to another AIU within the same CPC drawer, without disturbing client operations.

The Telum II cache structure is virtualized; the physical level of shared cache is implemented in a very robust SRAM wipe out correction and sparing design that benefits all virtualization levels. The cache allocation is fluid, dynamically changing the cache size and shape as needed, benefitting processor cores, Capacity Back Up, and On/Off Capacity on Demand. The L2 cache SRAM arrays now include additional sparing capacity for z17, improving tolerance for hard faults and reducing the need for complex repairs.

New for z17 is an I/O Data Processing Unit (DPU). The DPU moves functionality off the ASICs on the I/O adapters and in-boards it into the Telum II processor. The DPU hardware includes strong data protection including ECC on the microcontroller cores and the L2 interface, parity on data and control paths, and memory isolation for individual LCHIDs. The DPU has hierarchical error recovery designed to minimize the impact of faults, restore full I/O functionality, and avoid service actions as much as possible.

The z17 includes full, end-to-end Time Domain Reflectometry support for the processor drawer-to-processor drawer, DCM-to-DCM, and system Memory busses. In the event of a hardware failure on one of these busses, built-in circuitry can isolate and identify the faulty element(s) at the time of the error, simplifying and speeding up repairs.

The processor drawer-to-processor drawer bus has been redesigned for z17. All bits in the two CP interfaces now go across their own SMP cable. This bus has fault tolerance for wire failures, and the SMP cables can be replaced concurrently if multiple faults occur. An added benefit of concurrent cable repair is the reduction in the use of concurrent drawer repair.

The processor cooling system has been revised to a be a propylene-water solution based internal closed-loop system. In doing so, system installation and service has

been simplified and the need for a fill and drain tool as well as client storage of spare coolant has been eliminated. This improvement will also greatly reduce the time needed to prepare a machine for discontinuation, since draining the coolant is unnecessary. The z16 lift tool and service ladder can be reused for z17. This reduces storage requirements for clients with servers on both platforms. Remote Code Load, again available on z17, reduces the need for IBM and client service personnel to be onsite during firmware upgrades.

EMC Conformance

IBM z17 meets Class-A emissions and immunity standards for ITE equipment in various geographies around the world. Warning statements for these geographies can be found in the system's Installation Manual for Physical Planning (IMPP) which will be available at announce. The z17 IMPP is included in the Publications section of this announcement.

Hardware requirements

The hardware requirements for the IBM Z features and functions are identified. A new driver level is required. HMC 2.17.0 plus Machine Change Levels (MCLs) and the Support Element 2.17.0 will be made available on June 18, 2025. Clients should review software requirements for minimum MCLs and software PTF levels before IPLing operating systems.

HMC 2.17.0 Supported CPCs

Machine family	Machine type	Firmware driver	SE version
IBM z17	9175	61	2.17.0
IBM z16	3931, 3932	51	2.16.0
IBM z15	8561, 8562	41	2.15.0

Software requirements

IBM z17 requires at a minimum:

z/OS
– z/OS 3.2 with PTFs
– z/OS 3.1 with PTFs
– z/OS 2.5 with PTFs

The Abstract for the z/OS Upgrade Workflow will be updated on announcement day with the new Hardware server information. Clients can use that to find out how to

acquire the z/OS z17 Upgrade Workflow. This is an interactive z/OSMF Workflow which runs on the system. It contains all the information needed for z/OS to run on z17.

The z/OS Communications Server, a high availability enterprise transaction and data server, provides common applications, such as FTP, Telnet and the remote execution of applications. Built for optimum productivity, it provides a secure platform for developing and sharing mainframe workloads.

z/VM

– z/VM® 7.4 with PTFs
– z/VM 7.3 with PTFs

z/VM Advantage

IBM z/VM provides clients with a premier IBM Z Hypervisor for their existing workloads while also supporting a modernization journey to hybrid cloud, hosting enterprise-class virtual servers to exploit the advantages of IBM Z and IBM® LinuxONE in scalability, performance, high availability, and security. In conjunction with IBM z17, z/VM exploits the higher bandwidth and lower latency provided with Network Express adapters, including transparent VSwitch EQDIO support and NETH guest enablement. Performance DataPump support is enhanced with a power metrics dashboard that allows clients to monitor power consumption by a z/VM LPAR and by individual guests.

For z/VM support for the IBM z17 Family refer to the z/VM service required for the IBM z17 family servers on the z/VM product page.

21st Century Software VSE$^{n®}$ (21 CS VSEn®)

– 21st Century Software VSEn® 6.3 with PTFs
– IBM z/VSE® 6.2 will not support the IBM z17 Family as announced in Statement of direction: IBM z/VSE 6.2 support for the IBM z16

z/TPF

– z/TPF 1.1 with PTFs

Linux on IBM Z

Planned minimum distributions (all with service):

– SUSE SLES 15 SP6
– SUSE SLES 12 SP5
– Red Hat RHEL 9.4
– Red Hat RHEL 8.10
– Canonical Ubuntu 24.04 LTS
– Canonical Ubuntu 22.04 LTS

For minimum required and recommended distribution levels, see the Linux on IBM Z and IBM® LinuxONE tested platforms website.

Publications

The following publications are **available now** under IBM Systems Hardware Documentation:

Publication title	Order number
9175 Installation Manual for Physical Planning (IMPP)	GC28-7049
PR/SM™ Planning Guide	SB10-7184
IOCP User's Guide for ICP IOCP	SB10-7183
Planning for Fiber Optic Links (FICON/FCP, Coupling Links, OSA, and IBM zHyperLink Express)	GA23-1410
Hardware Management Console and Support Element Network Information Worksheet	SC28-7057

The following publications are **shipped with the product** and will be available on the planned availability date:

Publication title	Order number
9175 Installation Manual	GC28-7050
9175 Safety Inspection	GC28-7048
Installation Manual for TKE Workstations	GC28-7053
Systems Safety Notices	G229-9054
IBM Important Notices	G229-9056
Statement of Limited Warranty Part 3 - Warranty Information	GC28-7047
License Agreement for Machine Code	SC28-6872
License Agreement for Machine Code Addendum for Cryptography	GC27-2635
Systems Environmental Notices and User Guide	Z125-5823

The following publications **will be available on the planned availability date** under IBM Systems Hardware Documentation:

Publication title	Order number
9175 Service Guide	GC28-7051
9175 Parts Catalog	GC28-7052
Service Guide for Trusted Key Management (TKE) Workstations	GC28-7054
Service Guide for 2461 Hardware Management Console	GC28-7055
Service Guide for 2461 Support Element	GC28-7056
SNMP Application Programming Interfaces	SB10-7185
Capacity on Demand User's Guide	SC28-7058
Connectivity Mapping Tool User's Guide	GC28-7058
Hardware Management Console Web Services API (V2.17.0)	SC27-2646
IBM Dynamic Partition Manager (DPM) Guide	SB10-7188
Secure Service Container User's Guide	SC28-7062
Stand-Alone IOCP User's Guide	SB10-7186
FICON CTC Reference	SB10-7187
Maintenance Information for Fiber Optics (FICON/FCP, Coupling Links, OSA, and zHyperLink Express)	SY27-7698
Integrating IBM Remote Support into your Enterprise	SC28-7060
Hardware Management Console Security	SC28-7061
Remote Code Load for IBM Z Firmware	SC28-7068
IPL Machine Loader Messages	SC28-7063
OSA-Express client's Guide and Reference	SA22-7935
OSA/SF on the Hardware Management Console	SC14-7580

Publication title	Order number
OSA Integrated Console Controller User's Guide	SC27-9003

IBM Documentation

IBM Documentation provides a modernized user experience and makes it easier to find IBM product information for systems hardware, operating systems, and server software. Through a consistent framework, clients can efficiently find information and personalize their access. The IBM Z and IBM® LinuxONE publications are referenced under IBM Systems Hardware Documentation.

A list of deliverables for IBM z17 Model ME1 can be found on the Library Overview page within the IBM Documentation product page.

HMC and SE console documentation

At planned availability, the Hardware Management Console (HMC) and Support Element (SE) console documentation (Version 2.17.0) will be available from IBM System Hardware Documentation and the consoles.

Clients can also find HMC videos here: IBM Z Hardware Management Console Videos

IBM Redbooks®

Visit the IBM Redbooks site to find these z17 deliverables.

At announcement:	
IBM z17 Technical Introduction	SG24-8580-00
IBM z17 Technical Guide	SG24-8579-00
IBM Z Connectivity Handbook	SG24-5444-23
IBM Z Functional Matrix	REDP-5157-08
At General Availability (GA):	
IBM z17 Configuration Setup	SG24-8581-00
IBM Z Server Time Protocol Guide	SG24-8480-01

IBM Support

For installation and technical support information, visit IBM Support.

Terms and conditions

Products - terms and conditions

Warranty period

- **Warranty period** is one year.
- **Turkey only:** The warranty period is 2 years to comply with government requirements.

To obtain copies of the IBM Statement of Limited Warranty, contact your reseller or IBM. An IBM part or feature installed during the initial installation of an IBM machine is subject to the full warranty period specified by IBM. An IBM part or feature that replaces a previously installed part or feature assumes the remainder of the warranty period for the replaced part or feature. An IBM part or feature added to a machine without replacing a previously installed part or feature is subject to a full warranty. Unless specified otherwise, the warranty period, type of warranty service, and service level of a part or feature are the same as those for the machine in which it is installed.

Warranty Service

The specified level of maintenance service may not be available in all worldwide locations. Additional charges may apply outside IBM's normal service area. Contact your local IBM representative or your reseller for country and location specific information. IBM will repair the failing machine at your location and verify its operation. Clients must provide a suitable working area to allow disassembly and reassembly of the IBM machine. The area must be clean, well lit, and suitable for the purpose. The following service is available as warranty for your machine type.

- 24 hours per day, 7 days a week, same day response

International warranty service

International Warranty Service allows yo to relocate any machine that is eligible for International Warranty Service and receive continued warranty service in any country where the IBM machine is serviced. If you move your machine to a different country, you are required to report the machine information to your Business Partner or IBM representative.

The warranty service type and the service level provided in the servicing country may be different from that provided in the country in which the machine was purchased. Warranty service will be provided with the prevailing warranty service type and service level available for the eligible machine type in the servicing country, and the warranty period observed will be that of the country in which the machine was purchased.

The following information can be found on the IBM Support website:
– Machine warranty entitlement and eligibility
– Directory of contacts by country with technical support contact information
– Announcement Letters

Warranty service upgrades

The specified level of maintenance service may not be available in all worldwide locations. Additional charges may apply outside IBM's normal service area. Contact your local IBM representative or your reseller for country and location specific information.

IBM ON-SITE SERVICE: IBM will repair the failing machine at your location and verify its operation. You must provide a suitable working area to allow disassembly and reassembly of the IBM machine. The area must be clean, well lit, and suitable for the purpose.

General terms and conditions

Field installable features= Yes

Model conversions= Yes

Licensed Internal Code

IBM Licensed Internal Code (LIC) is licensed for use by a client on a specific machine, designated by serial number, under the terms and conditions of the IBM License Agreement for Machine Code, to enable a specific machine to function in accordance with its specifications, and only for the capacity authorized by IBM and acquired by the client. Clients You can obtain the agreement by contacting your IBM representative or visiting the License Agreement for Machine Code and Licensed Internal Code website.

Specific Machine Type Model: 9175-ME1

Planning information

Information on client responsibilities for site preparation can be found in the Installation Manual for Physical Planning (IMPP) on IBM Documentation website (see the link in the Publications section).

Machine Installation

The IBM z17 Multi Frame is a traditional, factory racked system. Installation is performed by IBM. IBM will install the machine in accordance with the IBM installation procedures for the Machine.

In the United States, contact IBM at 1-800-IBM-SERV (426-7378), in other countries contact the local IBM office.

Security, auditability and control

The IBM z17 uses the security and auditability features and functions of host hardware, host software, and application software.The client is responsible for evaluation, selection, and implementation of security features, administrative procedures, and appropriate controls in application systems and communications facilities.

Prices

Prices are subject to change without notice.

GST, QST, and sales taxes, where applicable, are extra.

For additional information on maintenance and pricing, contact your IBM Sales representative or your IBM Business Partner, or call 1-800-IBM-CALL (1-800-426-2255).

Appendix

Mainframe operating systems

Various sophisticated operating systems run on IBM mainframes—the security-rich, resilient and agile platform for integrating into your hybrid cloud strategy. IBM Z mainframes run on z/OS, Linux and z/TPF, with multiple operating systems (OS) often running on a single mainframe. IBM z/VM and KVM are hypervisors designed to run Linux, z/OS and z/TPF virtual machines (VM) as well as Red Hat OpenShift on IBM Z and IBM® LinuxONE servers. Each operating system and hypervisor have unique characteristics that bring the security, resiliency and agility of the mainframe platform to your hybrid cloud.

IBM Technology Lifecycle Services (TLS)

IBM Technology Lifecycle Services (TLS) leverages IBM watsonx AI, analytics-driven insights and automation to provide comprehensive, proactive and predictive support capabilities for IBM Infrastructure as well as leading third-party systems and software, through alliances and partnerships. As a single point of contact across the data center, TLS helps accelerate clients' AI and hybrid cloud journeys.

IBM Technology Expert Labs for IBM Z and IBM® LinuxONE

IBM Expert Labs is a professional services organization powered by an experienced team of product experts. This knowledgeable team brings deep technical expertise across software and infrastructure. Their skills include IBM data and AI, automation, sustainability, security, software-defined networking, IBM® Power®, IBM Storage, IBM Z and IBM® LinuxONE, IBM GDPS and IBM Cloud®. We use proven methodologies, practices and patterns to help our partners develop complex solutions, achieve better business outcomes and drive client adoption of IBM software, servers and storage.

IBM Z Global Skills Accelerator Program

The Global Skills Accelerator program combines prescriptive training, coaching, and on the job learning implemented at 140 employers skilling over 750 new IBM Z system administrators and application developers.

IBM Storage DS8000®

New IBM Storage DS8000 (10th Generation) is the latest innovation in enterprise-class storage for IBM Z and IBM i architectures. IBM Storage DS8000 helps your business achieve stronger data-driven decisions, access deeper insights and develop trust and confidence while increasing security and data privacy. IBM Storage DS8000 is the storage solution of enterprise data systems designed to ensure the availability of critical business workloads and lowering business risk of outages helps to ensure funding for future projects such as AI.

IBM Financing

IBM Financing provides finance solutions that enable organizations to efficiently support their IBM technology needs – helping them make the best decisions at the right time to ensure future growth.

- 40+ years' experience
- Global resources in 40+ countries
- Total solution financing across the IBM portfolio

IBM Consulting

IBM Mainframe Application Modernization Solutions accelerate transformation by integrating the IBM Z platform and an agile approach with open, portable applications and toolchains, that seamlessly aligns with hybrid-multicloud environments. We optimize applications and data on the platform that fits Client requirements and meets business objectives.

Sustainability of IBM Z

Learn more about what you can do to realize energy consumption and carbon footprint reductions by upgrading to the latest machine.