IBM z16

The digital transformation of business, institutions and society is accelerating faster than ever. A strong IT infrastructure is a component in business strategy, providing essential capabilities to deliver value and services in highly competitive markets in the digital economy.

IBM created the new IBM z16™ system to enable organizations to innovate and meet the needs of their customers and stakeholders with confidence and impact.

Built to build your business

The new core IBM Telum™ dual chip module has 16 processor cores, leverages the density and efficiency of 7nm chip technology, runs at 5.2 GHz or 4.6 GHz, and delivers increased performance and capacity across a wide range of workloads. There are up to 200 customer configurable cores on the Model A01 and 68 configurable cores on the Model A02 and Rack mount bundles.

The IBM z16 scales up to 200 configurable cores in a single model - the Model A01. The IBM z16 A01 is available with five options of core capacity – Max39, Max82, Max125, Max168 and Max200. The IBM z16 A02 is available with four options of core capacity- Max5, Max16, Max32 and Max 68.

The system offers up to 40 TB of Redundant Array of Independent Memory (RAIM) per system on the Model A01 and 16 TB on the Model A02 or it's Rack mount bundle equivalent.

IBM z16 single frame configuration

Highlights

- Integrate AI into business processes and IT operations to increase decision velocity and customer value. Protect data now and in the future with Quantum-safe cryptography
- Build a cyber resilient environment
- Modernize for hybrid cloud to deliver new value

- With IBM z16, process up to 300 billion inference operations/day with 1ms response time
- IBM z16 is the industry's first quantum safe system.
- Modernizing and running applications on IBM z16 is as little as 1/2 the OPEX of public clouds.

The Integrated Accelerator for IBM Z AI

The IBM z16 integrates new Artificial Intelligence (AI) acceleration via an on-chip AI coprocessor to reduce latency and deliver outstanding performance for in transaction inferencing.

Now organizations can embed AI directly into business processes and existing IBM Z® applications to improve business outcomes and deliver customer value in each interaction at unprecedented scale and speed within stringent SLA response time guidelines.

The Integrated Accelerator for IBM Z Sort

The IBM Integrated Accelerator for IBM Z Sort is a specialized processor on the IBM z16 core that helps to reduce the CPU costs and improve the elapsed time for eligible workloads.

IBM Z Sort is embedded in the CPU and memory latency is much lower than disk latency, sorting in memory is more efficient than sorting with both memory and disk workspace, resulting in educed elapsed time and CPU time.

The Integrated Accelerator for zEnterprise Data Compression (zEDC)

The Integrated Accelerator for zEDC function can reduce data storage and communications requirements and costs, as well as increase data transfer rates to boost throughput without adversely impacting response

times. The Integrated Accelerator for zEDC improves systems performance for pervasive encryption, so that customers can encrypt 100% of their data, 100% of the time.

Data Security for today and tomorrow

The IBM z16 represents a breakthrough in data security. Quantum-safe cryptography is embedded in the system to improve the resiliency to cyber-attacks from bad actors with future access to quantum computing resources. Today's cyber threats often involve harvesting encrypted data for decryption later when these resources can break today's encryption algorithms. The IBM z16 represents a step forward as customers have a safe and tested infrastructure that can deploy the more sophisticated and complex cryptography needed to protect today's sensitive data from cyber risks as they emerge.

IBM Z continues its leadership in data security with Quantum- safe security on the IBM z16, extending pervasive encryption, Fibre Channel Endpoint Security and other innovative advancements that make it the most securable computing platform in industry.

IBM Z Cyber Vault extends the cyber resiliency of the IBM z16 by enabling the quick recovery from ransomware attacks using EAL5 certified air-gap separation of compute and storage infrastructure and Safeguarded Copy to restore data using a protected and trusted copy of the data.

Flexible Infrastructure

Flexible Capacity for Cyber Resiliency on the IBM z16 enables customers to transfer capacity easily and efficiently between different data centers for disaster recovery, regulatory compliance, maintenance, and other business needs. Combined with IBM storage, GDPS® and System Recovery Boost, this function delivers an extremely high availability solution for mission critical workloads.

System Recovery Boost on the IBM z16 has been enhanced to include faster IBM middleware restart and faster SVC dump processing. System Recovery Boost enables customers to improve availability by accelerating the recovery time from planned and unplanned outages using all available processor resources during system shutdown and restart. It also provides the additional capacity to process workload backlog after the system has recovered from the outage.

Remote Code Load for IBM Z Firmware optimizes resilience and keeps your IBM Z system up to date with the latest features, fixes, and maintenance without requiring someone to be physically in the data center to install and monitor planned updates to your system. This optional feature provides secure, remote installation and monitoring by IBM for planned updates to your IBM Z.

The IBM z16 is built with a 19" frame that flexibly scales from one to four frames depending on the model and the configuration. IBM z16 Rack mount bundles enable clients to seamlessly integrate the system into existing racks with other equipment in a hybrid cloud data center. Customers migrating from the IBM z14® or earlier systems or consolidating Linux® workload from distributed systems will achieve significant sustainability improvements, including improved environmental efficiencies and floor space reduction for most customers. Changes to the footprint mean:

– The Central Processing Complex (CPC) drawer design has relocated the long-distance coupling to the PCIe+ I/O drawer.
– Support continues for both raised and non-raised floors as well as top and bottom exit I/O and power. All cabling is routed to the back of the frame with new brackets to contain cables.

– Customers have embraced intelligent power distribution unit (iPDU) technology implemented on all models of the IBM z16
– The doors are designed for acoustics and optimized for airflow. The IBM z16Multi frame model A01 requires 3-phase power, while the smaller Single frame (A02) and Rack mount configurations can use single-phase power.
– The use of the iPDU power may improve power efficiency and lower overall energy costs dependent on the required configuration.

The 19" frame technology supports the A3 operating class as defined by the American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE).1The benefit of having an A3 class rating is being able to save on Heat, Ventilation, and Air Conditioning (HVAC) costs due to the wide range of operating conditions that will now be available.

The optional IBM Z Hardware Management Appliance can be ordered with the IBM z16 to provide Hardware Management Console (HMC)/SE functions within the 19" frame, eliminating the need for a separate HMC outside of the server.

Continuous Compliance

The IBM z16 contains new capabilities to make compliance to PCI-DSS regulatory guidelines easier and more productive. Audit preparation times can be significantly reduced and require less staff to complete. The IBM z16 is integrated with the IBM Z Security and Compliance Center to monitor and record system, network and application data for changes and adherence to PCI-DSS standards. A user-friendly dashboard enables infrastructure personnel to easily and quickly generate reports that auditors need and to ensure a continuous compliance posture that mitigates the potential for noncompliance regulatory fines.

Tailored Fit Pricing for IBM Z enables price stability, particularly in an unpredictable hybrid cloud environment. The Tailored Fit Pricing Software and Hardware Consumption solutions are transformational pricing options for IBM z16. It provides simple, transparent, and predictable pricing for hardware and software when running the IBM z/OS® platform. The Tailored Fit Pricing options maximize cost predictability, while still supporting optimal response times and Service Level Agreements (SLA) compliance.

The Tailored Fit Pricing for IBM Z Hardware Consumption Solution provides instantaneous access to additional capacity for short, unpredictable spikes in business-critical z/OS workloads. It was designed to meet the demands of hybrid cloud workload on the platform. To meet these demands, the IBM z16 can now include, on top of the base capacity that you own, a fixed corridor of pay for use capacity. This always on corridor of consumption priced capacity will help you to alleviate the impact of short unpredictable spikes in workloads that are becoming more common in today's digital world.

IBM announced several open-source compilers to leverage the Integrated Accelerator for Z AI on the IBM z16 Telum processor chip, enabling programmers to embed inferencing easily and at scale in applications. Existing IBM compilers have been updated to exploit the latest IBM z16 architecture. This capability allows them to deliver cross-platform development and integration, operate with Java™, Swift or Node.JS and optimize enterprise workload performance without recompiling, as well as reduce the central processing unit (CPU) cycles needed to complete the job.

IBM z/OS Container Extensions (IBM zCX) capabilities for the z/OS environment allows developers to build and deploy Red Hat® OpenShift® containers or Docker workloads on z/OS. This can minimize the barrier to develop on the platform, while allowing the workloads to inherit the z/OS qualities of service benefits of high availability, integrated disaster recovery, scalability, workload manager, and integration with z/OS security.

The IBM z16 provides Linux deployments with more cores, more memory and cache innovations. The IBM z16 provides the platform for modernizing, developing, and managing on-premises, containerized applications. The new 19" frame opens new opportunities within a data center. A Linux platform can be deployed as a standalone server, or side-by-side with z/OS or z/VSE® environments on a single physical server. This allows for easy integration of Linux® workloads on the IBM z16 resulting in infrastructure benefits from tight data and application colocation, fast internal communications, and integrated high availability.

Fast and secure access to data

High-speed connectivity to data is critical to achieve balanced performance with storage device and exceptional transaction throughput. The IBM z16 offers:

– A 2-port FICON Express32S adapter that connects your IBM z16 to switches, directors, and storage devices at up to 32 Gbps. With support for native FICON®, High Performance FICON for IBM Z (zHPF) and Fibre Channel Protocol (FCP), the adapter helps meet the low latency and increased bandwidth demands of applications. When the Fibre Channel connection endpoints use the FICON Express 32S adapter or FICON Express16SA adapters to the IBM DS8900F storage, authentication of the endpoints is enabled.
– A set of OSA-Express7S 1.2 adapters that meet the increased networking bandwidth demands driven by high-speed processors and faster network-attached storage devices.
– Support for IBM zHyperLink™ 1.1, a direct connect, short distance, I/O adapter offering extremely low latency connectivity to FICON storage systems. The IBM Washington Systems Center offers the zBNA tool to help determine workload candidates that can benefit from this adapter.
– Shared memory communications that are used for either direct memory placement of data within the IBM z16, or host-to-host memory communications, using Remote Direct Memory Access over Converged Ethernet (RoCE) Express adapters, without significant TCP/IP processing costs.
– Support memory-memory communications is available for both z/OS and Linux on IBM Z.
– Support for IBM Internal Coupling Adapter Short Reach (ICA SR)1.1 and Coupling Express2 LR (long range) coupling interconnects with significant changes to the coupling limits with increase for ICA SR physical coupling links and coupling channel path identifiers (CHPID).

Configurations

Choose from three IBM z16 configurations	Multiframe IBM z16 A01 This configuration is designed for general-purpose use, offering a balance of performance, scalability and security suitable for a wide range of applications.	Single frame IBM z16 A02 This configuration is optimized for enhanced modularity and is available in a traditional factory-frame setup. It provides advanced capabilities and is ideal for environments requiring a robust and scalable solution within an industry-standard footprint.	Rack mount IBM z16 AGZ This rack-mounted configuration allows for installation in existing data center rack infrastructures, providing flexibility and integration with current systems. It combines IBM Z's strengths with the convenience of modern rack-mounted hardware.
Specifications			
Maximum number of engines	200	68	68
Maximum number of drawers	4	2	2
Maximum number of IO drawers	12	3	3
Number of frames	1 to 4	1	No frame
Colocate with storage/switch	No	No	Yes
Frequency	5.2 GHz	Specifications	4.6 GHz
Telum chip	Yes	Yes	Yes
Maximum memory	40 TB	16 TB	16 TB
Sizes	39, 82, 125, 168 and 200	5, 16, 32 and 68	5, 16, 32 and 68
Resources			
Specification sheets	Multiframe specifications	Single frame specifications	Rack mount Specifications
Energy efficiency	Multiframe carbon footprint	Single frame carbon footprint	Rack mount carbon footprint
Technical guides	Multiframe guide	Single frame guide	Rack mount guide
Interactive tour	Multiframe tour	Single frame tour	Rack mount tour

*Prices shown are indicative, may vary by country, exclude any applicable taxes and duties, and are subject to product offering availability in a locale.

Why IBM?

Position for today and tomorrow

Success in the digital economy is contingent on making IT a creator of value internal and externally. Fundamental to this is a flexible infrastructure that positions organizations strategically, leveraging AI and hybrid cloud, while protecting existing investments and improving sustainability.

The new IBM z16 delivers this with scale, agility, resiliency, performance, a security-rich environment, and a lower overall Total Cost of Ownership. The IBM z16 provides confidence in meeting the future, in a world of uncertainty.

For more information

IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: ibm.com/financing

1. Performance result is extrapolated from IBM internal tests running local inference operations in an IBM z16 LPAR with 48 IFLs and 128 GB memory on Ubuntu 20.04 (SMT mode) using a synthetic credit card fraud detection model (https://github.com/IBM/ai-on-z-fraud-detection) exploiting the Integrated Accelerator for AI. The benchmark was running with 8 parallel threads each pinned to the first core of a different chip. The lscpu command was used to identify the core-chip topology. A batch size of 128 inference operations was used. Results were also reproduced using a z/OS V2R4 LPAR with 24 CPs and 256GB memory on IBM z16. The same credit card fraud detection model was used. The benchmark was executed with a single thread performing inference operations. A batch size of 128 inference operations was used. Results may vary.

2. IBM z16 with the Crypto Express 8S card provides quantum- safe APIs providing access to quantum- safe algorithms which have been selected as finalists during the PQC standardization process conducted by NIST. https://csrc.nist.gov/Projects/post-quantum-cryptography/round-3-submissions. 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: https://www.etsi.org/ technologies/quantum-safe- cryptography." 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.

3. The total cost of ownership and operating expense comparison illustrates the cost benefits of Tailored Fit Pricing for IBM Z that can be achieved versus public cloud pricing structures for new and modernized enterprise applications. The enterprise application in each scenario entails deploying a Java-based application and a relational database. On z/OS, the Java application used WebSphere® and Db2® deployed on IBM z15. Each of the three public cloud provider scenarios included a software stack comprised of a commercial database, database and systems management tools, and WebSphere Hybrid Edition on the cloud. The workload is expected to grow 5% annually and is assumed to require 50% production capacity at a remote facility to comply with government disaster recovery regulations. Additional detail is published in Tailored Fit Pricing for IBM Z Offers Cost Effective Workload Growth.
Compared to Three Public Cloud Examples, May 2021: https://www.ibm.com/downloads/cas/XAR2QOBD