IBM Power Platform
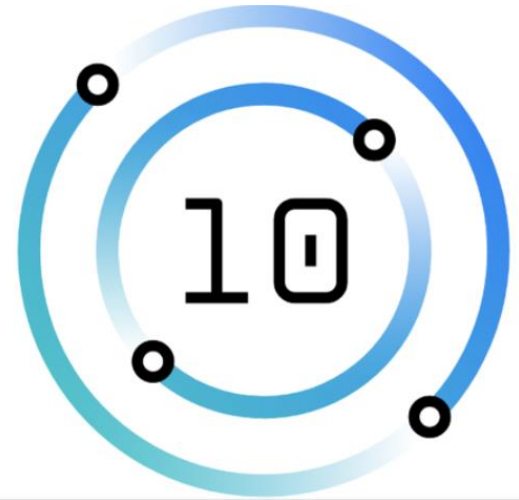
From Core to Cloud to Edge:
Introducing the new S1012 System

Fabian Michel
Senior Power & Storage Technical Specialist
IBM Technology, Belgium
fabian_Michel@be.ibm.com

in fabianmichel

May 2024

Generative AI is creating an inflection point in business transformation...

...enterprises that are more **intentional with hybrid cloud** will accelerate impact.

71%

of execs think its difficult to realize the full potential of a digital transformation without having a solid hybrid cloud strategy in place[1]

(1) IBM Transformation Index: State of Cloud

The 2023 CEO Study: Top priorities and challenges on the road to digital-first

Top Priorities

Productivity
48%

Tech Modernization
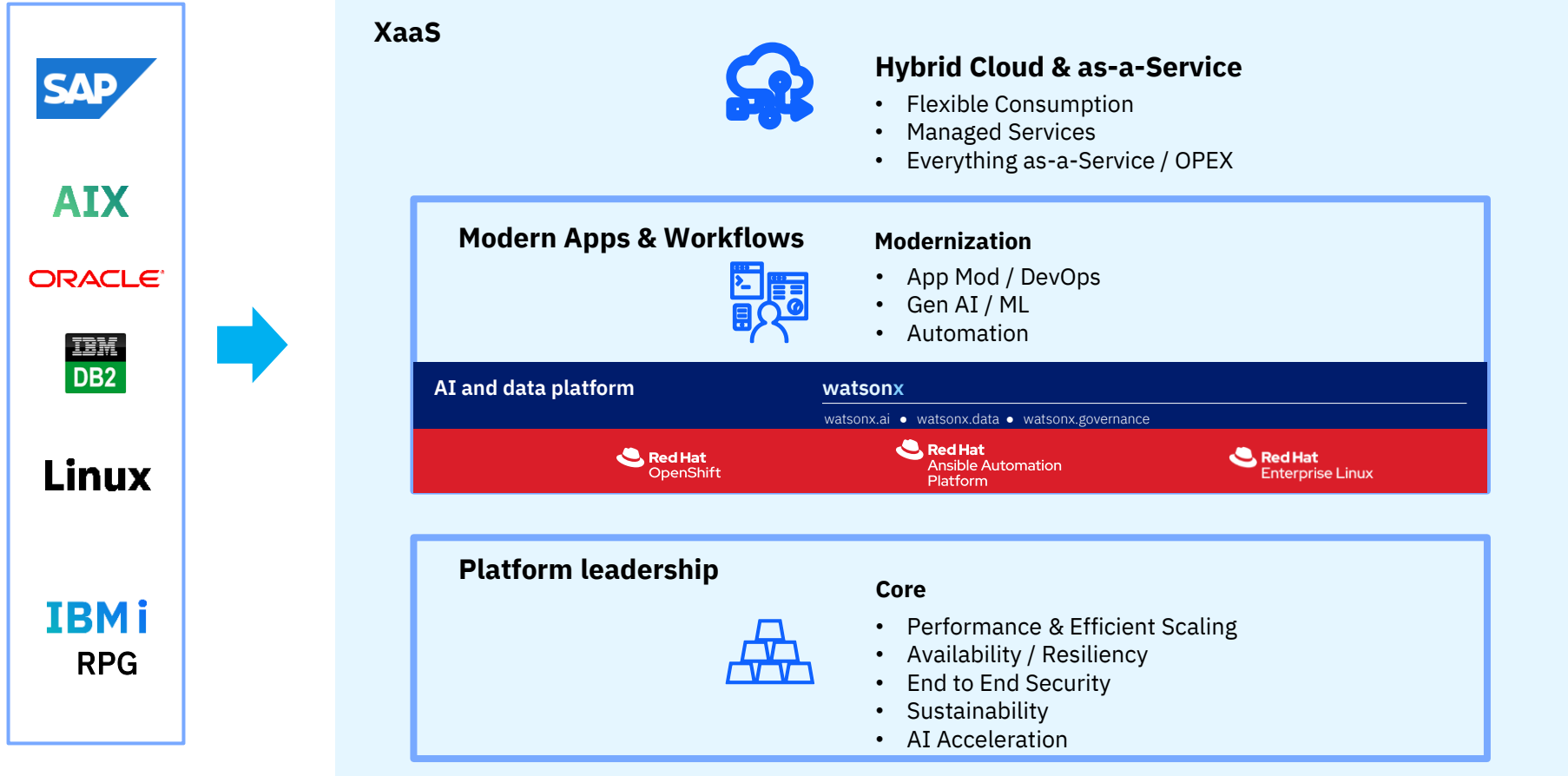# 45%

Top Challenges

Sustainability
42%

Cybersecurity
32%

IBM Power: Accelerate digital transformation with Hybrid Cloud and AI

SAP

AIX

ORACLE

IBM DB2

Linux

IBM i
RPG

XaaS

Hybrid Cloud & as-a-Service
- Flexible Consumption
- Managed Services
- Everything as-a-Service / OPEX

Modern Apps & Workflows

Modernization
- App Mod / DevOps
- Gen AI / ML
- Automation

AI and data platform   **watsonx**
watsonx.ai • watsonx.data • watsonx.governance

Red Hat OpenShift   **Red Hat Ansible Automation Platform**   **Red Hat Enterprise Linux**

Platform leadership

Core
- Performance & Efficient Scaling
- Availability / Resiliency
- End to End Security
- Sustainability
- AI Acceleration

# IBM Power: Infrastructure built for flexible and agile business transformation

IBM owned & managed infrastructure

PowerVS

- Public and Private (coming soon)

- VM-as-a-Service

- Dedicated host

- SAP RISE or IaaS

- Backup as a Service

- Production, Dev/Test, HA/DR

Customer owned & managed Datacenter/Colo

- Leadership platform portfolio for mission critical applications & data and consolidation

- Power Private Cloud with Dynamic Capacity and Flexible Consumption

Edge

New
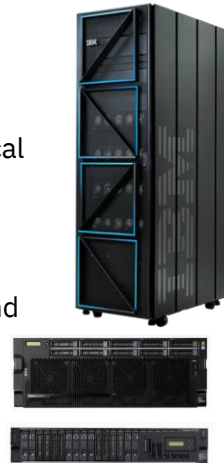
- Enterprise capability in the near edge (Security, Availability, AI acceleration)

- SMB, remote office, manufacturing, logistics, retail

IBM®

Delivering Cloud-to-Edge Excellence on Scale-Out Systems

New

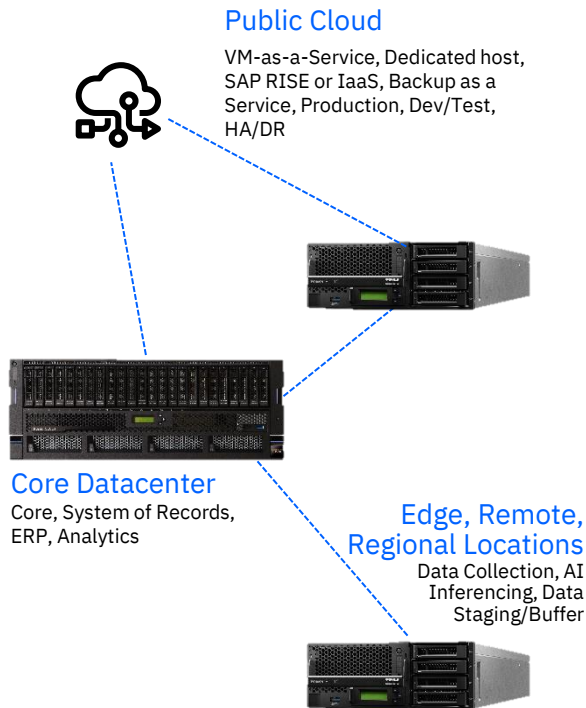| Edge Computing	Mainstream Computing		Core Computing		Cloud Computing
• Up to 8 cores and 256 GB memory footprints • Reduced IT footprint • Fully active, static servers	• Up to 16 cores and 2TB memory footprints • Up to 32% perf/price increase vs. P9 • Fully active, static servers • Lower technical requirements and competitive deals		• Up to 48 cores and 8TB memory footprints • Up to 1.4x system performance vs. P9 • Dynamic Capacity consumption with CUoD and PEP2.0 • Value-driven solutions and higher technical standards		• Unlimited cores and memory • Dynamic Capacity • Computing aaS
S1012 9028-21B	**S1014** 9105-41B	**S1022s** 9105-22B	**S1022 & L1022** 9105-22A / 9786-22H	**S1024 & L1024** 9105-42A / 9786-42H	**PowerVS**
• 1-socket, 2U, half-wide • Up to 8 cores per system • 1, 4, 8 SMT8 cores/socket • 4 ISDIMM slots • 256 GB memory • 4 PCIe HHHL slots (All Gen5) • 4 NVMe U.2 bays • Max of 6.4 TB of internal storage • Optional Internal RDX Media Bay • Rack and Tower form factors	• 1-socket, 4U • Up to 24 cores per system • 4, 8, 24 SMT8 cores/socket • 8 DDIMM slots • 1 TB memory • 5 PCIe FHHL slots (4 Gen5) • 16 NVMe U.2 bays • Max of 102.4 TB of internal storage • Optional Internal RDX Media Bay • Rack and Tower form factors	• 1,2-socket, 2U • Up to 16 cores per system • 4, 8 SMT8 cores/socket • 16 DDIMM slots • 2 TB memory • 10 PCIe HHHL slots (8 Gen5) • 8 NVMe U.2 bays • Max of 51.2 TB of internal storage	• 1,2-socket, 2U • Up to 40 cores per system • 12, 16, 20 SMT8 cores/socket • 32 DDIMM slots • 4 TB memory • 10 PCIe HHHL slots (8 Gen5) • 8 NVMe U.2 bays • Max of 51.2 TB of internal storage • L1022: max 25% of cores with other OS	• 1,2-socket, 4U • Up to 48 cores per system • 12, 16, 24 SMT8 cores/socket • 32 DDIMM slots • 8 TB memory • 10 PCIe FHHL slots (8 Gen5) • 16 U.2 NVMe bays • Max of 102.4 TB of internal storage • Optional Internal RDX Media Bay • L1024: max 25% of cores with other OS	• Delivered aaS off-premise • Unlimited cores • Unlimited memory • Unlimited storage and I/O • Add-on backup services including VTL

AIX **IBM i** **Linux** Red Hat OpenShift

All **Power VM** based servers

Introducing the IBM Power S1012 System

S1012 Overview

- New base system design to explore an alternate approach toward both existing and new use cases, with a Cloud-Core-Edge holistic infrastructure approach

- Dual system density unlocks possibilities for architecture best practices approach without compromising on space.

- 1-socket, 2U height, half-rack width, equipped with a 1, 4, or 8-core eSCM module, up to 256GB of memory, up to 6.4TB internal storage, 4 PCIe slots, and optional RDX.

- Available in rack-mounted option utilizing a bare-metal sliding drawer, and deskside-mounted option

- 3-year warranty systems with available value-added support services through Power Expert Care offerings.

Cloud-Core-Edge Infrastructure Approach

Public Cloud
VM-as-a-Service, Dedicated host, SAP RISE or IaaS, Backup as a Service, Production, Dev/Test, HA/DR

Core Datacenter
Core, System of Records, ERP, Analytics

Edge, Remote, Regional Locations
Data Collection, AI Inferencing, Data Staging/Buffer

Main Use Cases and Client Value

IBM i P05 & P10

- Security for mission-critical data running on IBM i
- Reduce IT footprint in ROBO/back-office
- Attractive price point comparable to Power9
- Improved SLAs to business area/local operations through an optimized HA footprint

Small Apps and DBs

- Increase availability and performance: up to 3x more performance per core vs. P8.
- Lower concerns with data privacy and data loss
- Support app modernization in small spaces through SNO with up to 32 threads per node

Cloud-to-Near Edge & AI

- Specialized AI silicon for inferencing operations w/o GPUs, suitable for Small Language Models and Vision Models
- End-to-end data protection through virtual and physical security features
- Embedded, enhanced remote management capabilities
- Sturdy, compact design fits near-edge locations
- Continuous offline availability without operation shutdown vs. cloud-based edge offerings

New IBM Power S1012 System

- New base system design to explore an alternate approach toward both existing and new use cases.

- It is a one-socket, 2U height, half-rack width, equipped with a 1, 4, or 8-core eSCM module, up to 256GB of memory, up to 6.4TB internal storage, 4 PCIe slots, and optional RDX.

- It is offered in two form factors:

  - A rack-mounted option utilizing a bare-metal sliding drawer.

  - A deskside-mounted option that is 54% smaller than S1014.

- 3-year warranty systems with available value-added support services through Power Expert Care offerings.

- Dual system density unlocks possibilities for architecture best practices approach without compromising on space

Main Use Cases for S1012

IBM i P05 & P10

- Security for mission-critical data running on IBM i
- Reduce IT footprint in ROBO/back-office
- Attractive price point comparable to Power9
- Improved SLAs to business area/local operation

Small Apps and DBs

- Increase availability and performance: up to 2.5x more performance per core vs. P8
- Lower concerns with data privacy and data loss
- Support app modernization in small spaces

Cloud-to-Near Edge & AI

- Built-in AI acceleration w/o purchasing, maintaining, or noise disturbance of GPUs
- Enhanced data security and remote management capabilities
- Data Lifecycle Management
- Manufacturing equipment management with IoT
- Suitable for Small Language Models and Vision Models

Worry-Free End-to-End Security

- Security is of paramount importance when systems are deployed in edge locations, beyond the confines of an enterprise datacenter. This is where the S1012 system excels, offering robust security measures to safeguard your operations.

- The S1012 system innovates by delivering all the IBM Power enterprise-grade security features in compact, powerful form factors.

- A well-thought design with security capabilities to continuously protect platform integrity across the main processor, service processor, and peripherals, and to address both virtual and physical security threats.

Online, Virtual Security

Transparent Memory Encryption
- All in-memory & in-transit data between memory and processor is encrypted
- Fully homomorphic encryption innovation: transactions are computed with active encryption, without the risk exposure and performance degradation of a de-encrypt process
- Works by design: no need for additional management setup, no performance impact or CPU cycle consumed, and no application recoding needed

Platform Integrity
- Secure boot of host and BMC
- Power10 enhanced CPU FSP/BMC isolation
- Performance-enhanced side-channel avoidance
- Power10 Return Oriented Programming (ROP) protection

At-Rest Data Protection
- Full disk data protection against theft with password-only access and encryption capabilities

Offline, Physical Security

Tower Door Lock
- Tower deskside systems come with a lock system to prevent unauthorized access
- Op-panel located outside the locked door to avoid unnecessary direct access to the system and the RDX drive

Rack Enclosure Protection
- The rack-mounted enclosure adds a protection layer to internal system components from external intervention.
- The enclosure offers a spring latch lock feature to reduce the risks of system theft.

USB Disable Option
- USB ports can be firmware disabled to prevent unauthorized direct connection

Hands-Free Remote System Management

- When systems are deployed in edge locations, typically, there is little to no skilled IT people nearby, which adds time, cost and complexity to the system management tasks.

- The S1012 system innovates by delivering native, enhanced remote management capabilities and possibilities.

HMC provides:

- Ability to configure and run multiple partitions, with or without VIOS
- Remote power-on of the system
- Shared remote console
- Guidance through concurrent maintenance operations
- Performance metrics
- And more...

+

New eBMC ASMI Remote Management Features

- No need to access and use a physical control panel to perform the most common configuration operations:
    - Select Alternate restart device
    - Select Load source device
    - Select Console adapter slot

- Improved remote management for IBM i partitions without HMC
    - Access DST without pressing buttons on the control panel
    - Initiate a Mainstore Dump
    - Display debug SRCs for error situations

IT Infrastructure Optimization

- Focused on an architecture best practices approach without compromising on space
- Remote Management, HA/Recovery, Single-Node Openshift, Watsonx.ai, Tape Hybrid Cloud Backup, Dev/Test, and upcoming use cases.
- Cloud-Core-Edge holistic infrastructure approach

Improve Remote IT management capabilities:

Workload   vHMC

2U

Improved, advanced high availability:

VMRM or PowerHA

Workload   HA Ready System

2U

Public Cloud
VM-as-a-Service, Dedicated host, SAP RISE or IaaS, Backup as a Service, Production, Dev/Test, HA/DR

Core

Datacenter
Core, System of Records, ERP, Analytics

Edge, Remote, Regional Locations
Data Collection, AI Inferencing, Data Staging/Buffer
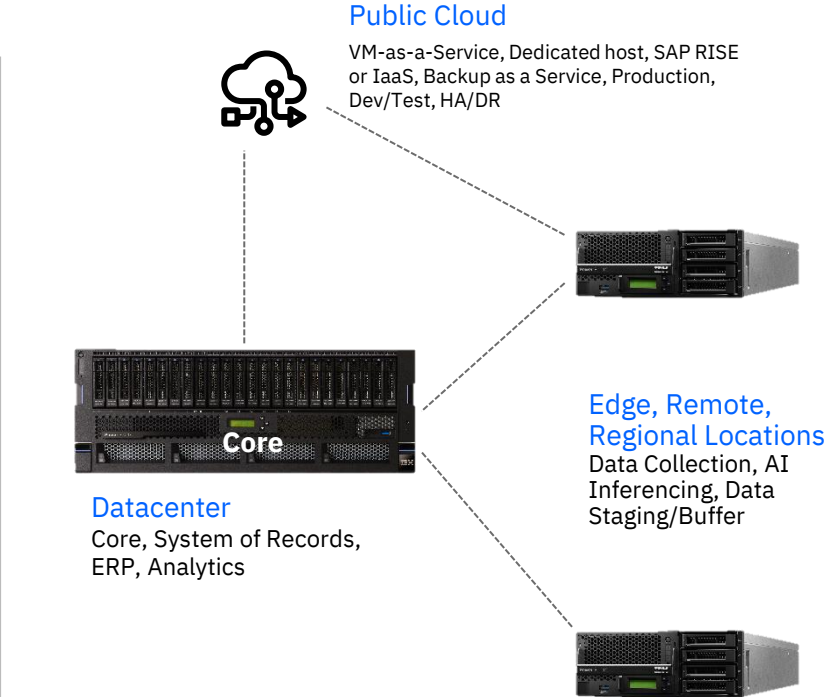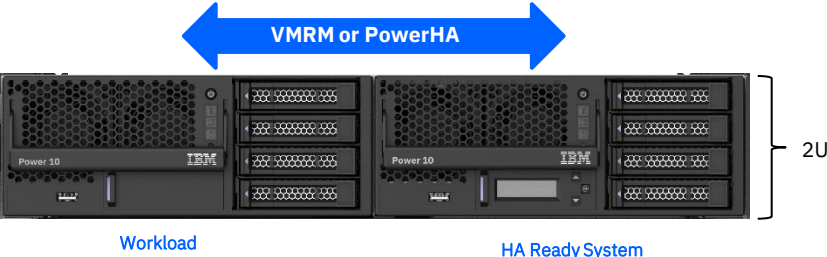
High-Availability Use Case Example

S1012 system design enables architecture best practices approach without compromising on space or budget. Up to 75% space optimization vs. 1S4U form factors in a dual-system installation.
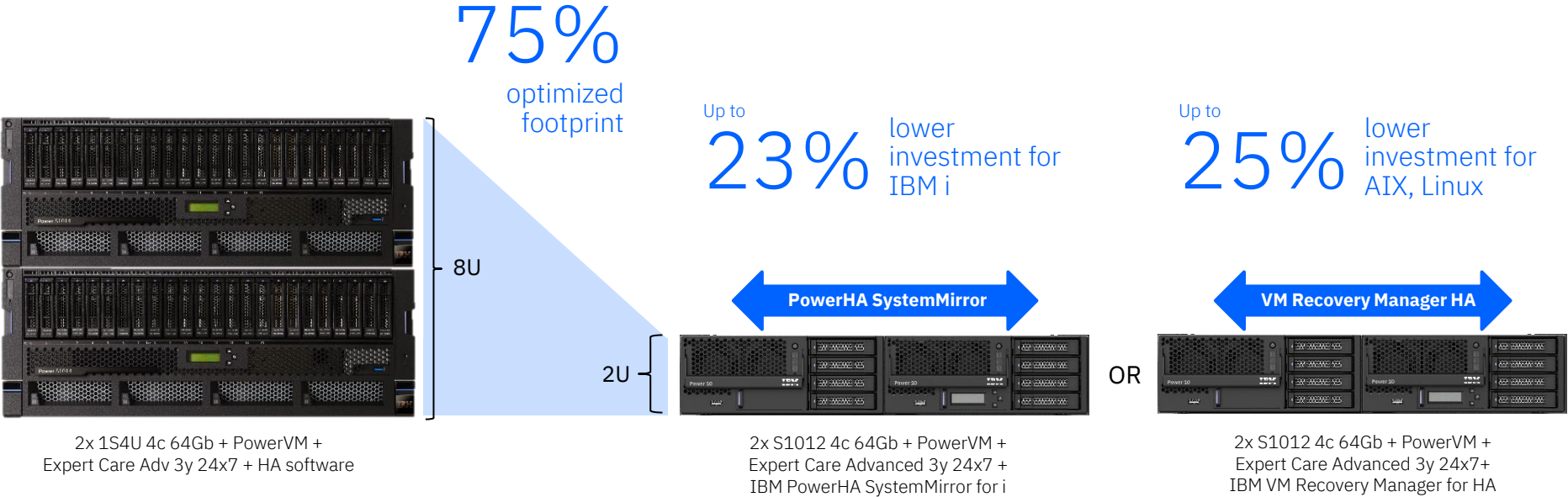
75%
optimized footprint

8U

2U

Up to
23% lower investment for IBM i

PowerHA SystemMirror

Up to
25% lower investment for AIX, Linux

VM Recovery Manager HA

OR

2x 1S4U 4c 64Gb + PowerVM +
Expert Care Adv 3y 24x7 + HA software

2x S1012 4c 64Gb + PowerVM +
Expert Care Advanced 3y 24x7 +
IBM PowerHA SystemMirror for i

2x S1012 4c 64Gb + PowerVM +
Expert Care Advanced 3y 24x7+
IBM VM Recovery Manager for HA

Expanding the Power Platform to Edge Computing

What use cases are businesses pursuing at the Edge?

- Predictive Security and Surveillance
- In-store Point-of-sales customer experience (scoring, face recognition, fraud detection)
- In-store real-time inventory management
- Distribution Center and Logistics Management
- Healthcare patient monitoring (support for HIPAA standards)
- Manufacturing equipment management with IoT
- Utilities: power grid optimization, reduce energy waste
- Telecom tower maintenance and data traffic optimization
- Crops, cattle, or wildlife protection and improvement of agricultural yields

Why are companies computing data at the Edge?

- As data continues to proliferate beyond datacenter environments, infrastructure tailored for edge optimization and AI will emerge accordingly
- Reduce network traffic to centralized systems avoiding bottlenecks and unnecessary cost
- Run AI inferencing workloads in ROBO and back-office locations outside mainstream datacenter facilities
- Advanced performance and availability SLAs for inferencing computing
- Secure the data and insights in and out of the AI models running locally
- Run inferencing of Small Language Models and Vision Models

Why IBM Power S1012 is the best option for Near-Edge deployments?

- Sturdy, compact design: fits near-edge locations such as remote offices, stores, and manufacturing

- Security: end-to-end data protection through virtual and physical security features

- Reliable: the rack-mounted and deskside enclosures provide additional protection from external threats

- Efficient Management: enhanced built-in remote management capabilities with a new eBMC console compatible with Redfish

- Modernization: support for Red Hat Openshift and planned support for Single-Node Openshift

- AI - enabled: Specialized AI silicon for inferencing operations w/o GPU cost and noise

- Online: Full offline availability without operation compromise

Public Cloud
Back Up/DR, Dev/Test

Edge, Remote, Regional Locations
Data Collection, AI Inferencing, Data Staging/Buffer

Core Datacenter
Core, System of Records, ERP, Analytics
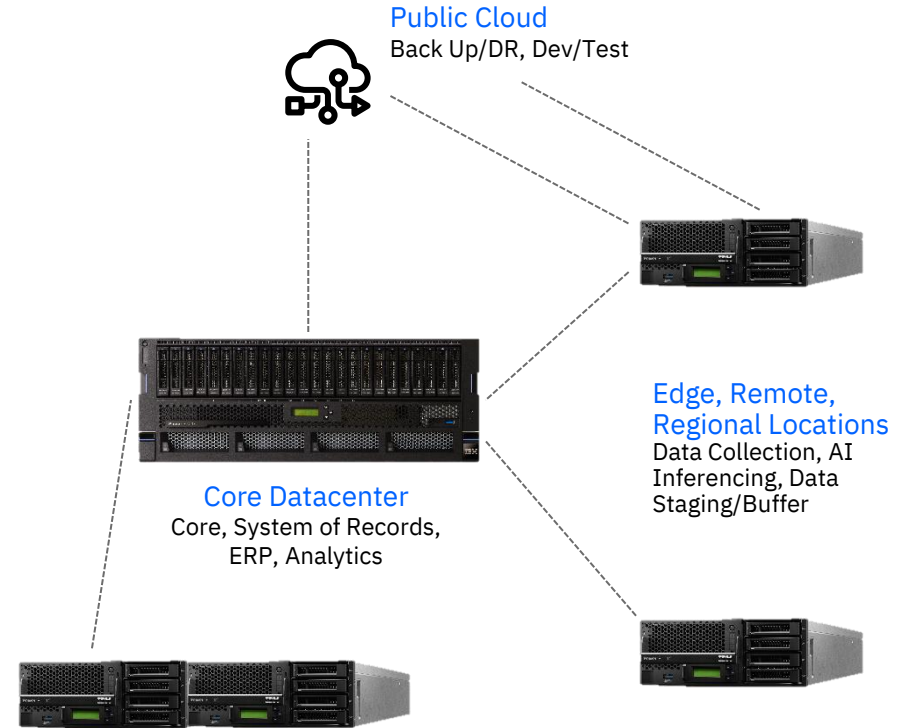
End to end security with full stack encryption

Stay ahead of current and future threats with support for:
- **Quantum-safe cryptography**
- **Fully homomorphic encryption**

Applications
Hyper-sensitive data

Databases
Sensitive in-use, in-flight and at-rest data

File and data sets (AIX EFS)
Sensitive data tied to access control for in-transit and at-rest data

Memory encryption
All data in memory

Full disk and tape
(AIX LV encryption, IBM i ASP encryption)
Protect at-rest data

APP

DB

DATA SET / FILE

DISK TAPE

DATA
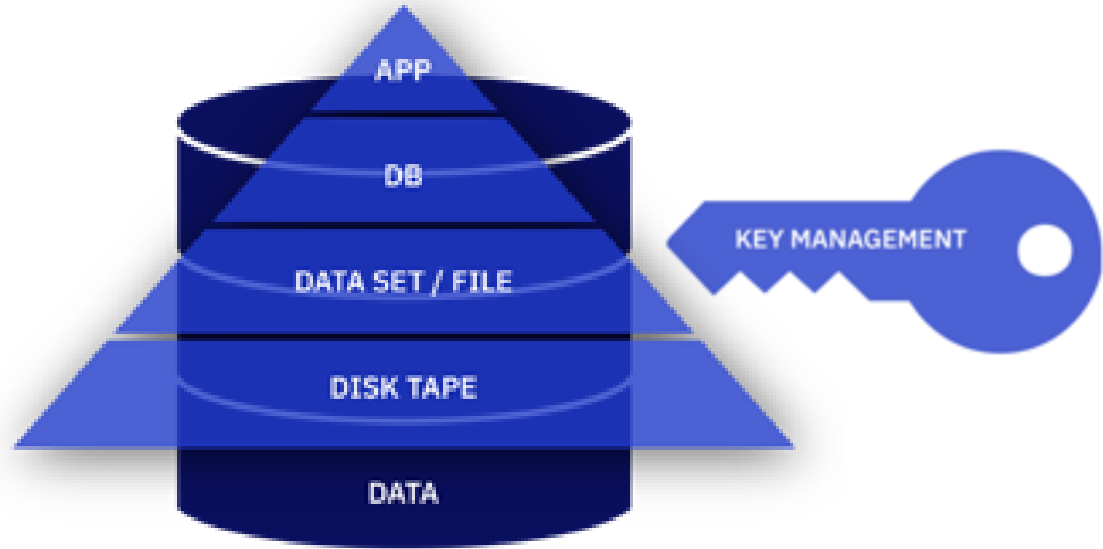
KEY MANAGEMENT

Transparent memory encryption with:
- No additional management setup
- No performance impact or CPU cycle consumed
- No application recoding needed

Blazing fast hardware-accelerated encryption
- 4x crypto engines in every core

Data Fabric instead of data silos

Streamline insights and automation with
In-core AI inferencing and machine learning

MMA: Matrix Math Accelerator

+ Extreme Bandwidth

- Enable the incorporation of AI into Enterprise workloads while maintaining Enterprise qualities of service

- Seamlessly integrate AI for improved business insights

- Scalable solutions for Deep Learning inference

- Broad support for Machine-Learning and Analytic workloads

Built-in AI acceleration debunks the need for purchasing, maintaining, increasing noise disturbance of GPUs
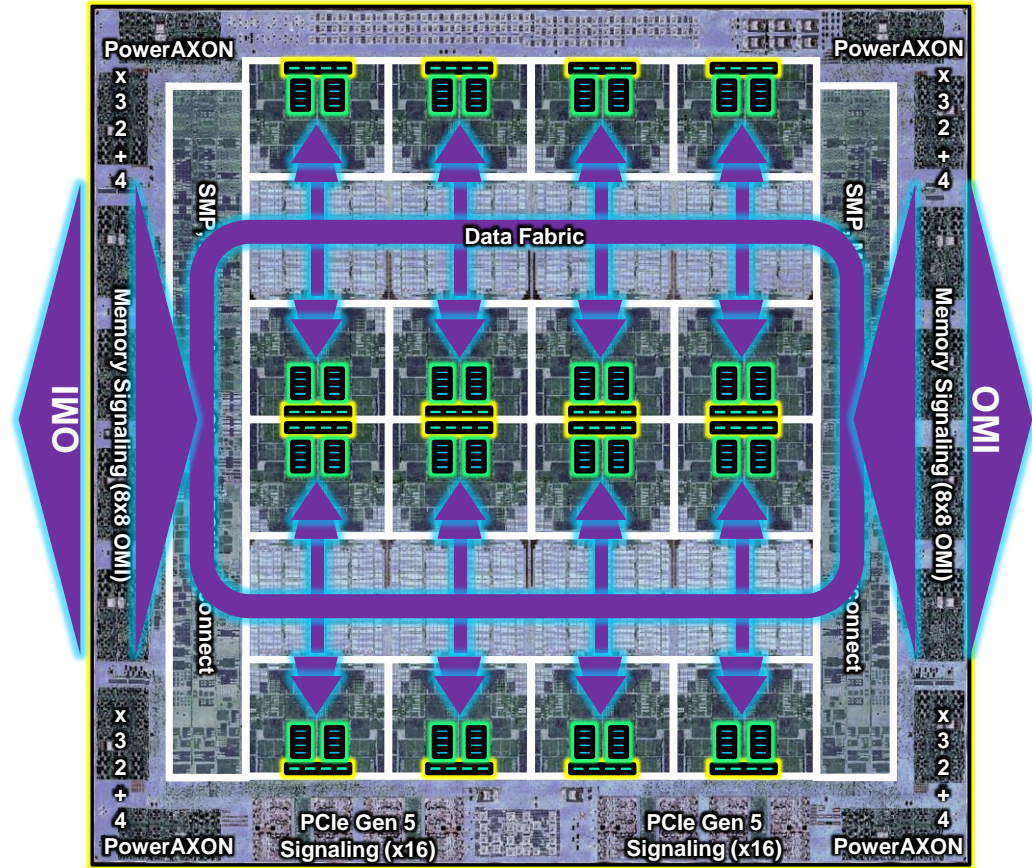
Power10 has AI acceleration in every core

- 4 MMA engines per core
- 8 SIMD engines per core
- High bandwidth data-path

Flexibility on Initial Order Configurations

The S1012 introduces an innovative, compact, but flexible design with powerful, secure, and reliable capabilities. It is offered in two form factors and shipped in different set-ups that best suit your customers:

Rack-mounted Options

Filler + S1012

- For customers with requirements fully covered by one S1012 system in the first months or year, with a ready-to-go expansion provisioned
- The 2nd bay of the enclosure comes with a filler to avoid dust and inaccurate airflow through the rack.

S1012 + S1012

- For customers looking to adopt infrastructure best practices such as high availability, advanced remote management
- For customers looking to expand the IBM Power footprint beyond its current workload, i.e., with a focus on AI and container-based apps, open-source databases

RDX + S1012

- For customers with a backup architecture relying on Tapes, the 2nd bay of the enclosure can be filled with an RDX drive.
- Although it sits externally to the system, it provides an internal RDX-like performance through a direct-attach connection at the back of the enclosure

S1012 Tower

- For customers that require reliable and secure computing capacity in places without racking or datacenter-like infrastructure
- The tower desk-side enclosure fits one S1012 system and an optional RDX Tape Drive, with a door lock function and an outside op-panel, dust filter, and a low, office-level acoustics certification.

Rack-to-Tower and Tower-to-Rack conversions are supported when the 4-core processor module is configured. The 1-core processor module is supported only on the Tower form factor. The 8-core processor module is supported only on the rack-mounted form factor

System Comparisons

Features	S914	S1022s	S1014	S1012
GA	3Q20	3Q22	3Q22	2Q24
MTM	9009-41G	9105-22B	9105-41B	9028-21B
Sockets	1 SCM	2 eSCMs	1 eSCM	1 eSCM
Socket Power Max	190W max	240W max	415W (r24c), 240W (rack), 195W (tower)	240W (rack), 195W (tower)
Module Core Counts	8 (rack only), 6, 4	8, 4	24[1], 8 (rack only), 4	8 (rack only), 4, 1 (tower only)
Core Count (SMT-8) Max per system	8	16	24	8
Memory Slots	16 DDR4 ISDIMMs	16 DDR4/DDR5 DDIMMs	8 DDR4/DDR5 DDIMMs	4 DDR4 ISDIMMs, memory buffers down on planar
Memory capacity	1 TB	2 TB	1 TB	256 GB
Memory bandwidth	170 GB/s	409 GB/s	204 GB/s	102 GB/s
System PCIe slots	10 PCIe slots 4 G5 x8 or G4 x16 direct 4 G5 x8 direct 2 G4 x8 direct	10 PCIe slots 4 G5 x8 or G4 x16 direct 4 G5 x8 direct 2 G4 x8 direct	5 PCIe slots 1 G5 x8 or G4 x16 direct 3 G5 x8 direct 1 G4 x8 direct	4 PCIe slots 2x G5 x8 or G4 x16 direct 1x G5 x8 direct
Slots for Internal Storage	Dedicated	General	General	Dedicated
Drives	18 SAS HDD/SDD or 4 NVMe U.2 Optional RDX	8 NVMe U.2	16 NVMe U.2 Optional RDX	4 NVMe U.2 Optional RDX
PCIe4 expansion drawer	0.5 w/ LPC Cable Adapter	2 w/ LP Cable Adapter	0.5 w/ LPC Cable Adapter	N/A
Non-Supported Functions (available on other P10 systems)				LPM*, vPMEM, HNV, vNIC

* Planned for later availability

IBM Power Expert Care Service Tiers for S1012

- Below are the support services that will be available on IBM Power S1012

	Warranty	Basic	Advanced	Premium
IBM Hardware Maintenance	3-year 9x5, next business day, IBM onsite limited	3, 4, 5-year 9x5, next business day, IBM onsite repair	3, 4,5-year 24x7, same business day, IBM onsite repair	3, 4,5-year 24x7, same business day, IBM onsite repair
IBM Software Maintenance	3-year Software Support Services	- To match selected year term of Power Expert Care	- To match selected year term of Power Expert Care	- To match selected year term of Power Expert Care
Predictive Alerts	- 9x5 Alerts through Call Home Cloud Connect - 9x5 call back from IBM Representatives	- 9x5 Alerts through Call Home Cloud Connect - 9x5 call back from IBM Representatives	- 24x7 Alerts through Call Home Cloud Connect - 24x7 call back from IBM Representatives	- 24x7 Alerts through Call Home Cloud Connect - 24x7 call back from IBM Representatives
Enhanced Response Time				- Part of Premium tier
Technical Account Manager (TAM)				- Part of Premium tier
Remote Code Load				- Part of Premium tier

Back-Up

Performance Benchmarks for S1012

The S1012 performance is at parity or slightly better than the S1014 benchmarks due to incremental hardware, firmware, and IBM i software performance optimizations available after the Power10 announcement timeframe when the S1014 scores were measured and published.

rPerf

			S1012	
			8c	4c
			234.2	130.4
S1014	8c	254.2	0.92	-
S1014	4c	129.4	-	1.01
S914	8c	183.2	1.28	-
S914	4c	80.9	-	1.61
S814	8c	91.8	2.55	-
S814	4c	63.0	-	2.07

2.5X 8-core S1012 vs 8-core S814

1.6X 4-core S1012 vs 4-core S914

CPW

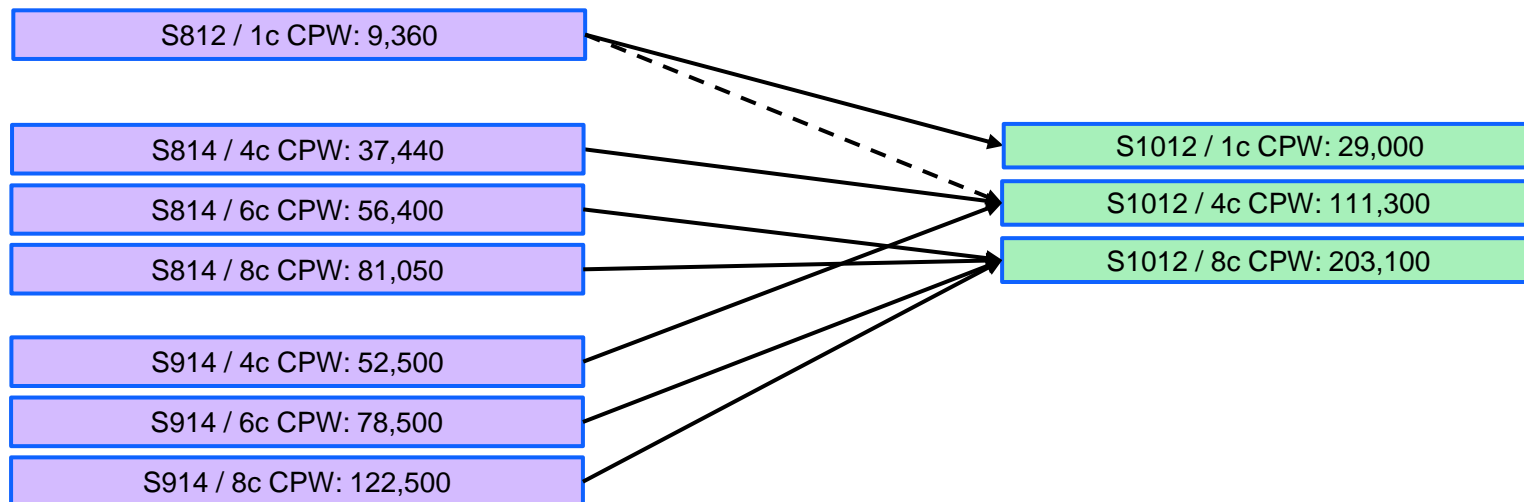|			S1012		
			8c	4c	1c
			203,100	111,300	29,000
S1014	8c	198,200	1.02	-	-
S1014	4c	106,300	-	1.05	-
S914	8c	122,500	1.66	-	-
S914	4c	52,500	-	2.12	-
S922	1c	19,000	-	-	1.53
S814	8c	81,050	2.51	-	-
S814	4c	37,440	-	2.97	-
S812	1c	9,360	-	-	3.10

3X 4-core S1012 vs 4-core S814

3X 1-core S1012 vs 1-core S812

2X 4-core S1012 vs 4-core S914

IBM IBM i Migration Options / Technology Refresh

IBM i

- Consolidate two 4C or 6C systems onto a single S1012
  - Reduce data center footprint and energy consumption by up to 75%
  - Reduce core counts

- Consolidate two 4-core systems onto a single S1012
  - Reduce data center footprint and energy consumption by 75%
  - Reduce core counts

| S812 / 1c CPW: 9,360 |

| S814 / 4c CPW: 37,440 |
| S814 / 6c CPW: 56,400 |
| S814 / 8c CPW: 81,050 |

| S914 / 4c CPW: 52,500 |
| S914 / 6c CPW: 78,500 |
| S914 / 8c CPW: 122,500 |

| S1012 / 1c CPW: 29,000 |
| S1012 / 4c CPW: 111,300 |
| S1012 / 8c CPW: 203,100 |

IBM Power S1012 1-Pager

- **P10 eSCM Feature**
  - 1x 8 BC (Rack only)
  - 1x 4 BC
  - 1x 1 BC (IBM i only)
  - Max socket power 240W

- **OMI Memory with DDR4 ISDIMMs**
  - 4 OCMB+ 4 ISDIMMs
  - 16GB, 32GB, 64GB, 102 GB/s peak
  - Max 256 GB capacity

- **Internal Storage**
  - 4x NVMe Slots
    - 0.8 TB, 1.6 TB NVMe 15mm U.2
  - Max 18W
  - 6.4 TB capacity max
  - Optional RDX

- **PCIe HHHL Slots**
  - 2 PCIe G5 x8 / G4 x16 direct
  - 2 PCIe G5 x8 direct

- **Form factor**
  - Half-wide 2U rack
  - 26" system depth
  - Includes desk-side

- **Rack Shippable option**

MTM 9028-21B
GA June 2024

Rack and Desk-Side

AIX and VIOS
- AIX73D+: AIX Version 7.3 with the 7300-01 Technology Level and Service Pack 2 (P10 mode)
- AIX72Z: AIX Version 7.2 with the 7200-05 Technology Level and Service Pack 8 (P9-compat mode)
- VIOS 4.1.0.20
- VIOS 3.1.4.40 (P9 mode)

IBM i
- IBM i 7.4 TR10
- IBM i 7.5 TR4

Linux
- RHEL 9.2

- **OS**
  - IBM i
    - 1 core is native only
    - 4 & 8 core are native or virtual client
  - AIX
  - Linux
- **Hypervisor**
  - PowerVM
- **RAS**
  - Concurrent maintenance on NVMe
  - No concurrent maintenance on PCIe adapters
  - Redundant cooling, concurrent maintenance supported
  - Redundant power, concurrent maintenance supported
    - (1+1) 800W CRPS PS, 110 or 220 VAC
  - Customer setup, install & repair
- **Energy Efficiency**
  - 80+ Titanium Power Supply Compliant
  - EPA Energy Star Compliant
  - Built-in Advanced Thermal & Power Management
- **Enterprise BMC**
  - Op-panel (Base), LCD op-panel (Featured)
  - 1 1GbE, 1 USB/UPS 2.0 (all rear)
- **Host I/O**
  - 2 USB 3.0 (front, internal for RDX)
- **System Management (optional)**
  - HMC/vHMC (not redundant), Novalink, PowerVC (rack only)
- **Certifications**
  - FCC: Class A
  - Acoustics: General Business Category 2E for Desk Side and Data Center Category 1B for Rack
  - Environment: ASHRAE A3
    - 5-40C, 8-85% RH, 3050m max