

IBM Power S1012: Bringing Power10 to the Edge and to Smaller Power Environments

May 11 2024 | For IBM Business Partners, Technical Sellers, and Solution Architects

By Kevin Gee, Senior Technical Solution Architect

The IBM Power platform has traditionally been associated with large enterprise environments running AIX, IBM i, and Linux workloads where reliability, security, and operational consistency matter more than raw hardware cost. With the introduction of the IBM Power S1012, IBM is extending Power10 into a segment of the market that historically has been harder to address economically: smaller distributed environments, remote offices, and edge deployments.

For partners and architects looking for a practical entry point into the Power10 portfolio, the S1012 fills an important role. It brings the same Power10 architecture and platform characteristics into a system designed for customers that may not need a larger data center-class footprint.

What Is the S1012?

The IBM Power S1012 (MTM: 9028-21B) is a single-socket Power10 server available in either a 2U half-width rack configuration or a tower form factor. The platform supports 1, 4, or 8 SMT8 processor cores operating between 3.0 and 3.9 GHz, up to 256 GB of DDR4 memory across four DIMM slots, and as many as four NVMe U.2 drives for internal storage.

The system also supports dual redundant 800W Titanium-rated power supplies capable of operating on either standard 110V or 240V power. In practical terms, that makes the platform easier to deploy in locations where traditional data center infrastructure may not exist.

More importantly, the S1012 establishes the lowest acquisition entry point within the current Power portfolio. For customers that delayed Power10 adoption because of cost, space, or operational concerns, the S1012 changes the conversation considerably.

Comparing the S1012 against the S1014 and S1022

Understanding how the S1012 fits relative to the S1014 and S1022 is important because, although all three systems share the same Power10 architecture, they address very different operational requirements.

The Power S1014 (9105-41B) is a larger 4U system supporting up to 24 cores, up to 1 TB of memory, significantly greater internal storage density, and additional PCIe expansion capacity. Smaller 4-core and 8-core S1014 configurations also use the eSCM processor module, which means they retain the same single-socket licensing characteristics as the S1012.

In practice, the S1014 is usually the better fit when customers expect substantial growth in memory utilization, virtualization density, storage capacity, or I/O requirements over time. More often, customers are trying to minimize footprint, simplify deployment, or stay within a very constrained infrastructure budget. That is where the S1012 tends to fit well.

The S1022 (9105-22A) addresses a different class of workload entirely. It is a 1-to-2-socket, 2U full-width server using DCM (Dual-Chip Module) processor packaging, supporting up to 40 cores, as much as 4 TB of memory, larger virtualization footprints, and substantially greater expansion capability.

That DCM architecture also has software licensing implications that matter in customer discussions. Oracle and several other ISVs count each DCM as two sockets, which can materially increase licensing costs relative to eSCM-based systems such as the S1012 and smaller S1014 configurations.

As a result, the S1022 tends to align better with consolidated AIX environments, larger database workloads, OpenShift clusters, or dense virtualization deployments where core count and memory density matter more than footprint or entry cost.

Conversely, if a customer's workload comfortably fits within 8 cores and 256 GB of memory, the S1012 often delivers the same Power10 architecture with significantly lower hardware, infrastructure, and software licensing costs.

What tends to matter most operationally in distributed environments is the S1012's physical density. Two systems can be mounted side-by-side within a single 2U rack space. For retail sites, healthcare facilities, manufacturing floors, or remote office deployments where rack capacity is limited, that can simplify both deployment planning and ongoing infrastructure management.

Important S1012 Use Cases

1. IBM i Entry and Refresh Deployments

The most immediate fit for the S1012 is the IBM i customer base operating Power7, Power8, or early Power9 infrastructure in smaller environments.

Many of these organizations are not looking for large-scale consolidation platforms. They need a modern system with current security capabilities, strong per-core performance, manageable operating costs, and the operational reliability traditionally associated with the Power platform.

A single-core tower configuration can operate from standard office power without requiring specialized facilities infrastructure. For IBM i P05 environments, the economics are considerably more approachable than many earlier generations of entry Power systems.

2. Edge AI and Localized Inferencing

One of the more relevant newer use cases for the S1012 involves lightweight AI inferencing at the edge.

Each Power10 core includes Matrix Math Accelerators (MMAs), which provide built-in acceleration for certain AI and analytical workloads without requiring discrete GPU hardware. That creates opportunities to run lightweight inferencing workloads, including smaller language or vision models, directly within retail locations, healthcare environments, warehouses, or manufacturing facilities.

For organizations evaluating localized inferencing, this can help reduce infrastructure complexity in environments where deploying dedicated GPU-based systems may not be operationally practical.

Power10's transparent memory encryption capabilities also help protect data resident in memory without requiring application modification, which can be valuable in distributed environments with stricter security or compliance requirements.

3. Small Database Consolidation

The S1012 is also a practical fit for smaller database consolidation projects, particularly those involving Oracle Database Standard Edition 2.

Because Oracle SE2 licensing is socket-based rather than core-based, the S1012's eSCM design can help contain licensing costs while still providing substantial compute headroom for consolidation. That distinction becomes especially relevant when compared with DCM-based platforms such as the S1022, where Oracle licensing treatment can materially affect total solution cost.

For smaller deployments consolidating multiple application or departmental databases, the economics can be favorable.

4. Compact High Availability Designs

The physical density of the S1012 also enables compact high-availability configurations.

Two S1012 systems can occupy the same 2U rack space while supporting replication or clustering technologies such as IBM PowerHA SystemMirror, IBM i HA solutions, or IBM Virtual Machine Recovery Manager. In some environments, one system may host infrastructure services such as a virtual HMC while the adjacent system supports production workloads.

Neither the S1014 nor S1022 can provide the same HA density within a comparable footprint.

5. Small OpenShift Deployments

The S1012 can also serve as a compact platform for smaller OpenShift or containerized application deployments, particularly in edge or appliance-style environments where minimizing physical footprint is important.

Who Should Be Looking at the S1012?

The S1012 is particularly relevant for organizations that historically viewed Power as operationally larger or more expensive than their deployment requirements justified.

That includes:

- IBM i customers refreshing aging infrastructure
- Distributed retail and healthcare environments
- Manufacturing and logistics edge deployments
- Smaller AIX consolidation projects
- Remote office virtualization environments
- Customers evaluating lightweight AI inferencing and/or containers at the edge

For IBM Business Partners and system integrators, the S1012 fills a gap that existed in the Power portfolio for several years: a compact Power10 platform that can be deployed economically outside of a traditional data center while still preserving the reliability, security, and operational consistency associated with the platform.

For many customers that previously assumed Power was either too large or too expensive for smaller deployments, the S1012 provides a much more practical starting point.

Model	Max Cores	Cores per Processor Options	Min-Max Memory	Max VMs	Operating Systems	Max Internal Disk Capacity	System PCIe Slots	Form Factor
S1012	8	1, 4, 8	32 GB 256 GB	160	AIX, Linux, IBM i	6.4 TB (4 x 1.6TB NVMe U.2)	2 PCIe x8 Gen5 or x16 Gen4 2 PCIe x8 Gen5	2U, half-width Also a deskside option
S1014	8, 24 ¹	4, 8, 24 ¹	64 GB 1 TB	160	AIX, Linux, IBM i	102.4 TB (16 x 6.4TB NVMe U.2)	1 PCIe x16 Gen4 or Gen5 x8 3 PCIe x8 Gen5 1 PCIe x8 Gen4	4U Also a deskside option
S1022	40	12, 16, 20	64 GB 4 TB	800	AIX, Linux, IBM i	51.2 TB (8 x 6.4TB NVMe U.2)	4 PCIe x16 Gen4 or Gen5 x8 4 PCIe x8 Gen5 2 PCIe x8 Gen4	2U

- 1 - S1014 has a special 1-socket configuration with 24 cores, available only for Oracle configurations. Other configurations are limited to 4 or 8 cores.
- S1022 - IBM i instances are limited to 4 cores
- S1012 - IBM i instances are can run at 1 core
- COD available on the S1014 and S1022
- S1012 can run PowerVM.

The IBM Power S1012 will be available in July 2024. For technical specifications, configuration guidance, and ordering information, contact your IBM business partner, IBM representative, or visit <http://ibm.com/products/power-s1012>.

Kevin Gee is a Senior Technical Solution Architect and IBM Champion. He can be reached at geek@aixgeek.org.