Simple, Easy-to-Use, and Rapid Deployment of Cloud Infrastructure

Executive Summary

For Cloud Service Providers (CSPs) and Managed Service Providers (MSPs), the challenge is increasing the capacity, reliability, and performance of their cloud services while keeping their costs under control. The infrastructure supporting hybrid cloud services, multi-cloud services, and AI is being rapidly adapted to meet surging customer requirements in enterprise and SMB businesses.

Service providers must keep pace with this undeniable growth in cloud services and optimize their time to revenue. They must rapidly deploy a simple, scalable, and easy-to-use infrastructure. Customers depend on their CSPs and MSPs for cloud services that provide the applications these enterprise and SMB organizations rely upon to run their businesses.

As a leading manufacturer of servers and storage, Supermicro provides this holistic, agile, end-to-end infrastructure with rack-ready systems optimized for service providers of all sizes.

Cloud Infrastructure Deployments are Rapidly Growing

Driven by hybrid and multi-cloud deployments, the worldwide revenues for cloud service infrastructure are seeing double-digit growth each year—with no signs of slowing down in 2025-2026. Cloud services offer customers a strong value proposition, combining scalability, cost efficiency, flexibility, global reach, security, disaster recovery/business continuity, and technological advancements. Figure 1 depicts the key value provided by cloud services.

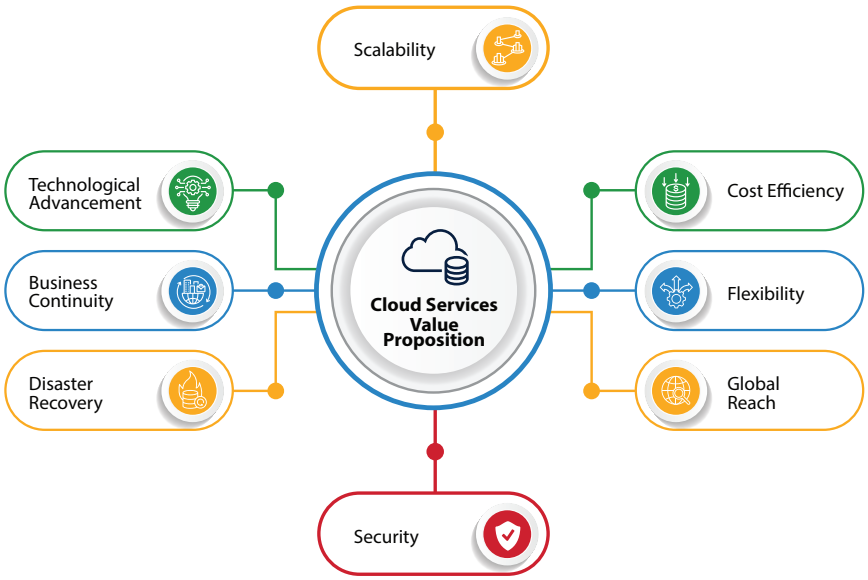

Figure 1: Cloud services provide a strong value proposition to customers across many industries

Many customers are accelerating the migration of their enterprise applications – deployed for years on scalable systems in the data center – and moving them to the hybrid cloud. This migration presents an excellent opportunity for CSPs and MSPs to capture more market share by making it easier and more cost-effective for their customers to migrate away from traditional on-premises IT infrastructure.

Evidence of business customers' move to cloud services is compelling: IDC's Worldwide Quarterly Enterprise Infrastructure Tracker[1] showed that customer spending on compute and storage infrastructure products for cloud deployments increased 18.5% in Q4 2023, compared with the year-ago period, rising to $31.8 billion ($USD). Shared cloud accounted for 44.9% of customers' total infrastructure spending this year. The outlook for 2025 is equally strong. The move to cloud services is undeniable.

Other market research firms are finding the same thing. Gartner[2] predicts that more than 70% of enterprises will use industry cloud platforms by 2027, up from 15% in 2023 – with GenAI adoption playing a vital role in increasing the rate of growth and Forrester Research[3] finds that a cloud-first model is taking hold at customer sites worldwide, from the Americas to EMEA to Asia/Pacific.

Cloud Infrastructure Deployment Challenges and Needs

A new generation of compute and storage infrastructure is being built and deployed by CSPs and MSPs to deliver today's applications and data and to do so with performance and quality. To be successful, CSPs and MSPs must overcome infrastructure deployment challenges to provide reliable, efficient cloud services to their customers. Figure 2 depicts these cloud infrastructure deployment challenges.

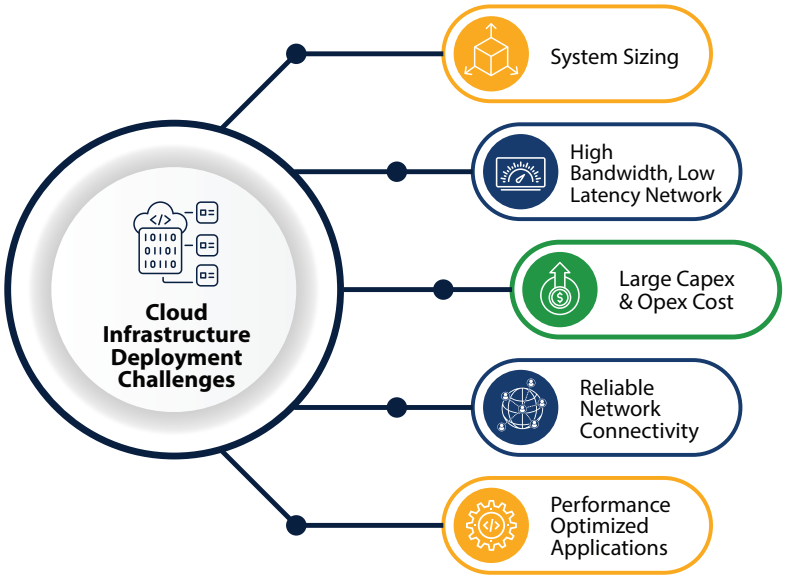

Figure 2: Key deployment challenges with cloud infrastructure

Hyperscalers, such as AWS, Azure, and Google Cloud, are known for a high degree of in-house customization for the systems they deploy in their very large-scale data centers. However, much of the cloud services infrastructure worldwide is being built for customers by a much wider group of cloud providers—including many hundreds of large, medium, and small service providers.

CSPs and MSPs deliver services to enterprises and businesses in many regions worldwide, adapting their services to support local geographic and customer requirements for software compatibility and security. Many focus on cloud services for enterprises and well-defined industry sectors, such as healthcare, financial services, retail, manufacturing, mining, and energy resources.

An essential part of Supermicro's services is guiding the customers, easing their transition to cloud services, and ensuring they are easy to learn and use. Further, with Supermicro's agile supply chain and rigorous testing, a wide range of CSPs and MSPs can quickly ramp up operations and resolve issues to ensure their clients experience minimal disruption and get maximum satisfaction.

When applied with governmental compliance and security, these capabilities allow these service providers to compete successfully to meet customers' business needs.

Simplicity in Design for Service Provider Platforms

CSPs must leverage entire racks of server and storage systems that can be put to work immediately as an integral part of scaling cloud services for their customers, supplying the services as demand grows.

CSPs and MSPs are looking to install and add full racks quickly and easily – matching growing demand and new requests from business customers. Special configurations slow things down and impact IT productivity for CSPs instead of enabling them to integrate the systems inside the rack-optimized data center infrastructure.

From a business perspective, the speed and quality of the services matter most – not climbing a learning curve about "how to" configure the server and storage systems that provide those services. Systems must be optimized for performance and scalability to ensure responsiveness for the services they deliver to end customers. These are Supermicro's key differentiators.

The Supermicro Difference

Supermicro has four themes for the differentiators its CSP customers see compared to other cloud infrastructure providers. Together, they speed up the rapid deployment of custom-designed and tested systems to fit within a CSP's power/cooling envelope inside an expanding data center (Figure 4).

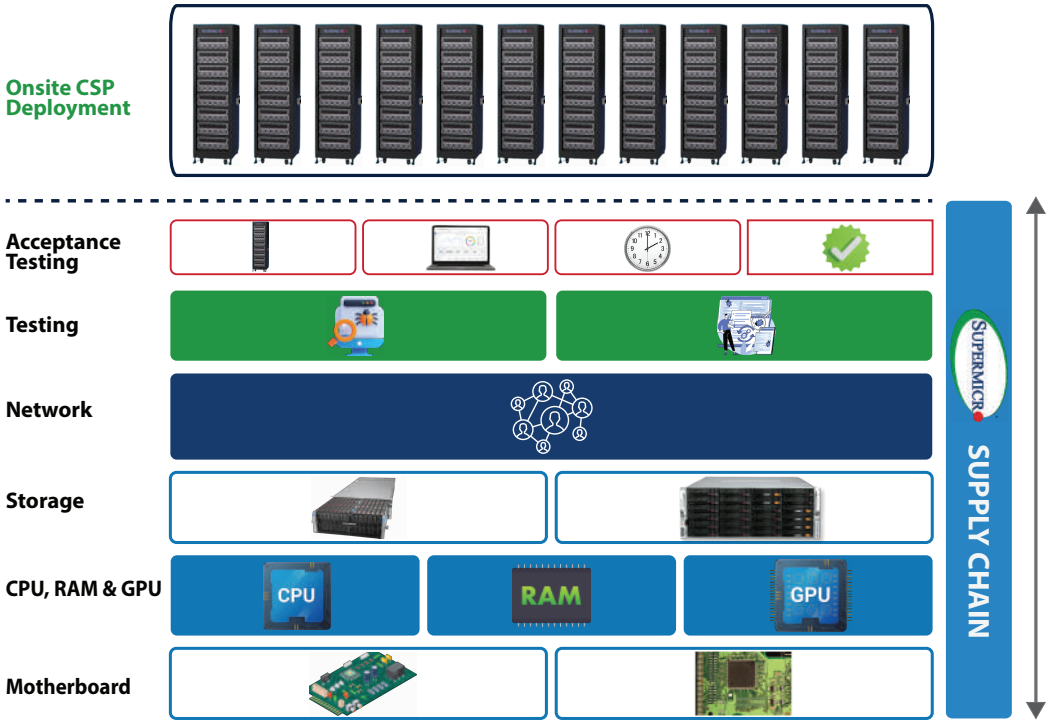

Figure 3: Agile supply chain with deep manufacturing partnerships speed up time-to-value for CSPs

Here are the four leading differentiators of Supermicro's buildout of ready-to-deploy rack-integrated systems, providing capacity and speed to meet and exceed CSP customers' growing businesses:

- ***Rack-readiness:*** The whole focus of providing CSP systems is preparing the rack units for quick delivery to the customer. This means that Supermicro systems can be used immediately, reducing the customer's time-to-market. From the ordering process to the configuration, burn-in, testing, and installation on-site, the entire process takes days, not weeks.
- ***Choices:*** Supermicro offers an extensive catalog of server and storage products that are available for inclusion in rack units. These choices give CSPs more flexibility when delivering racks that meet the capacity and performance needs of enterprise and SMB business customers.
- ***Speed of service:*** Supermicro has the architects, designers, developers, and personnel who will pull together all the systems, storage, and connectors to meet the CSP customer's ordering specifications. The company's expertise and its integration centers work closely together to complete and install orders on-site at the customer's location, as needed, to expand capacity and boost speed. It's all part of Supermicro's end-to-end service for its CSP customers.
- ***Power efficiency:*** Supermicro systems support expanded capabilities within power/cooling envelopes that comply with CSP customer data-center specifications, environmental policies, and standards. Power efficiency is a leading attribute for Supermicro rack deployments. It is another differentiator for Supermicro's integrated rack units, including strong support for liquid cooling and air-cooling that reduces power and heat requirements.

It's What's Inside (the Systems) that Counts

Ready-to-go systems carrying powerful CPUs and GPUs support a strong foundation of "building blocks" for competitive services delivery. It includes built-in support for microservices, generative AI (GenAI), and AI/ML cloud services within an enterprise, as well as finding the analytics data that will feed the customer's business and allow it to grow rapidly and with agility.

Full racks are built with high-performing engines for cloud services and AI. Indeed, a variety of "engines" are being deployed inside service providers' data centers for compute and storage. Systems based on AMD EPYC™ processors and AMD Instinct™ accelerators are powerful, high-performing engines for cloud services and AI.

These systems fit into standards-based data center racks, supporting network connectivity and interconnects that deliver high-speed data transfer between systems in industry-standard racks. This approach to ready-to-go infrastructure speeds up links between multiple racks, which is beneficial for CSP data centers with dozens, hundreds, or thousands of racks.

Systems must have adequate performance and capacity and comply with widely accepted networking and data management industry standards.

Business customers – the SMB and enterprise end-customers leveraging cloud services to run apps and data—look to CSPs and MSPs to provide those in-the-rack system building blocks. To meet their end-customer needs, CSP and MSP service providers should rely on systems and software providers with a proven track record regarding support for widely used networking switches and compliance with national and international interconnect standards.

Reliable network connectivity is crucial for cloud systems to communicate with each other and for users to access the applications. CSPs must ensure adequate bandwidth, low latency, and high availability of network connections to support cloud deployments, especially in geographically distributed environments.

To overcome critical deployment challenges for cloud infrastructure, CSPs and MSPs follow a set of best practices that help them overcome obstacles, speeding up deployment time and reducing time-to-revenue for themselves and the end customers accessing their services.

Best Practices for Rapid Deployments

Overall, CSPs and MSPs need to reduce the time to set up and speed up the delivery of cloud services to their business customers. They look for the simplicity of rack-optimized systems, support for key industry standards, and well-architected racks for consistent, efficient, and secure cloud services. Figure 4 depicts the best practices for CSPs.

Supermicro works closely with CSPs to determine their infrastructure requirements, configure CSP systems on a rack-unit basis for quick deployment, and provide service and support for the racks installed in CSP data centers to deliver cloud services.

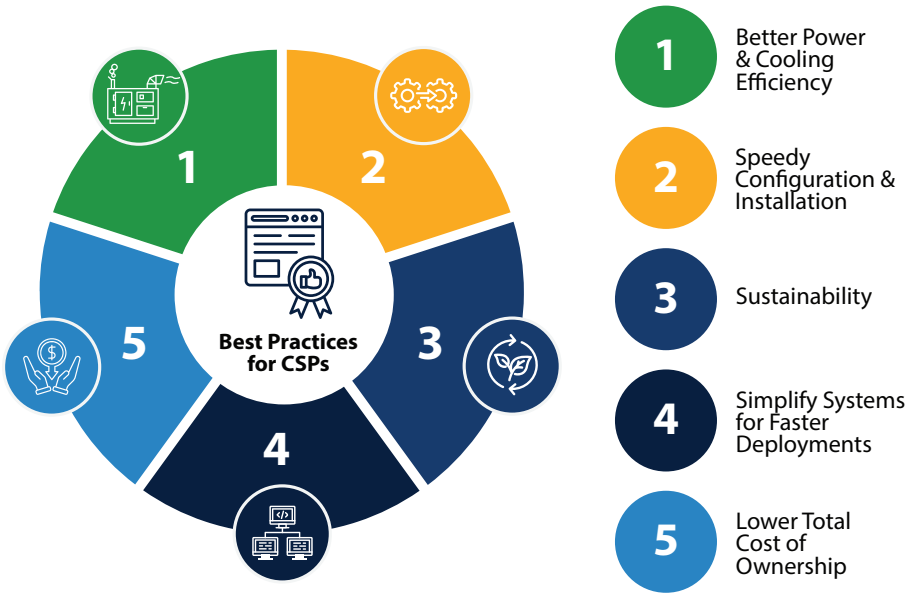

Figure 4: Best Practices for CSPs deploying cloud infrastructure

Overall, CSPs and MSPs need to reduce the time to set up and speed up the delivery of cloud services to their business customers. They look for the simplicity of rack-optimized systems, support for key industry standards, and well-architected racks for consistent, efficient, and secure cloud services. Figure 4 depicts the best practices for CSPs.

Supermicro works closely with CSPs to determine their infrastructure requirements, configure CSP systems on a rack-unit basis for quick deployment, and provide service and support for the racks installed in CSP data centers to deliver cloud services.

Systems designed and tested to be installed by the rack are best for rapidly deploying scale-up and scale-out cloud infrastructure. Rack-level integration benefits customers as they modernize their business applications, manage their daily business processes, and achieve agility by adapting quickly to ever-changing business conditions.

Best Practices for Cloud Services

- CSPs need to reduce the time to set up and speed up the delivery of cloud services to their business customers.
- CSPs need a partner with comprehensive global support to provide a total infrastructure solution – combining servers, storage, software, and services inside racks that lower a CSP's total cost of ownership (TCO).
- Now, CSPs can leverage rack-level integration and testing to reduce the time it takes to deliver cloud services to their end customers.
- This approach produces the most efficient, cost-effective CSP solutions across many applications.
- Rack-integrated units support a wide range of customer workloads, including IaaS (infrastructure as a service), PaaS (platform as a service), and SaaS (software as a service).
- SaaS is the largest segment of the cloud-enabled as-a-service marketplace – bringing enterprise and SMB applications to customers worldwide.

Figure 5 depicts the key considerations for deploying best practices.

Figure 5: Considerations to Deploy Best Practices for Cloud Infrastructure

Supermicro's AMD EPYC CPU Based Systems

Supermicro offers the H13 series of systems based on the latest AMD EPYC™ 9004 and 8004 series processors. The H13 systems deliver rack density, scalable performance, and energy efficiency for demanding, high-performance workloads.

Supermicro has introduced AMD EPYC™ powered H13 WIO, Hyper-U™ series, Petascale All-Flash Storage™, and a range of 1U and 2U Tower and 8U 8-node MicroCloud™ servers based on AMD EPYC™ processors.

Additionally, there are AMD-based systems that feature the new Zen 4c architecture and AMD 3D V-Cache™ Technology for Edge and Cloud uses.

The result is a series of workload-optimized servers for Edge, Cloud, AI, and HPC high-performance computing that CSPs can combine to meet each end customer's specific business requirements.[4]

AMD EPYC™ PROCESSORS

If organizations aren't careful, massive amounts of data can result in high energy use. Efficiency is key for data-intensive applications, and AMD EPYC™ processors power the most energy-efficient servers available.

In addition to delivering overall better performance per watt, AMD EPYC™ processors can closely match CPU resources with application requirements, creating even greater efficiency. For example, some analytic applications do not scale well to high core counts.

Using high-frequency AMD EPYC ™ processors can increase per-core performance to speed up these applications without the burden of carrying additional cores not essential to the mission. Some technical computing applications operate best when processors are equipped with large L3 caches. AMD EPYC ™ processors with AMD 3D V-Cache™ technology free CPUs to process data with fewer cache misses and therefore unimpeded performance.

Whether an enterprise needs as few as 8 cores or as many as 128 cores, or specialized processors, AMD EPYC ™ processors offer the freedom to choose.

All core features—including memory capacity, I/O bandwidth, and security features—are consistent within each processor family.

The Supermicro Advantage

Supermicro's deployment and support engineers and its speed and accuracy of full racks reduce the time and cost of system setup for CSPs. Supermicro's deep experience in building rack-optimized infrastructure allows CSPs to move more quickly when configuring, deploying, and maintaining racks for their business customers.

Based on thousands of customer deployments, Supermicro offers rack units filled with servers, storage, and software that support specific solutions for enterprise and SMB customers. These enterprise and SMB computing building blocks are proven, tested, and optimized for fast and efficient operations.

Having the right resources is essential to this rack-ready approach to cloud services infrastructure. Supermicro has four rack-integration sites worldwide: San Jose, CA; Malaysia, The Netherlands, and Taiwan. Each integration center contributes to the company's expertise in optimizing infrastructure systems that enable efficient cloud-ready services to thousands of CSP sites spanning all major geographies worldwide.

The integration work that unifies the systems inside each rack unit is done by Supermicro architects, engineers, and integration experts. CSPs do not have to do the integration and testing work; Supermicro does it. Because of this, building and testing racks for cloud services are simplified, tested, high-performing, and reliable.

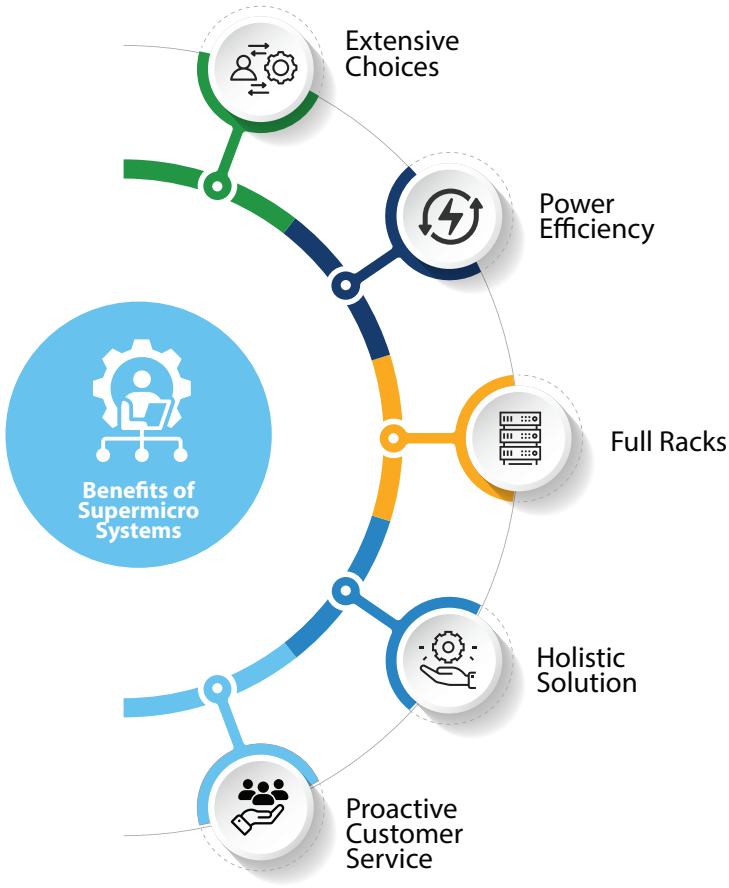

Figure 6: Benefits of Supermicro systems

Key benefits to CSP customers include:

- Supermicro reduces the costs of configuration and setup – while speeding the delivery of new and expanded services to enterprise business customers.
- Rapid deployment helps CSPs by speeding the delivery of new and expanded services to business customers.
- Supermicro's Rack Plug and Play Cloud Infrastructure™ is a recognized solution that speeds rack-level configuration and testing, responding to customers' requirements for high-performing infrastructure.[5]
- Integrated testing in Supermicro's integration centers reduces CSP customers' total cost of ownership (TCO), reducing the customers' time and materials needed to design, build, and deploy new systems.
- As a systems and storage provider, Supermicro provides, tests, and confirms the entire end-to-end infrastructure that delivers a CSP's cloud services.

For More Information

For more information on Supermicro solutions for CSPs, and how to get started with rack-enabled solutions, please visit our website (https://www.supermicro.com/en/products/aplus) or contact Supermicro AI experts directly to schedule a meeting.

---

[1]https://www.idc.com/getdoc.jsp?containerId=prUS52001524#:~:text=The%20dedicated%20cloud%20infrastructure%20segment,decline%201.4%25%20to%20%2457.6%20billion

[2]https://www.rtinsights.com/gartner-forecasts-robust-growth-in-worldwide-public-cloud-spending/#:~:text=Gartner%20predicts%20that%20over%2070,role%20in%20supporting%20this%20growth

[3]https://www.forrester.com/blogs/category/cloud-computing/

[4]Broad Range A+ Servers for Data Center, Cloud, AI | Supermicro

[5]https://www.supermicro.com/en/solutions/rack-plug-and-play